PERCEPTUAL LEARNING OF DYSARTHRIC SPEECH

A thesis submitted as fulfilment of the requirements for the degree
of Doctor of Philosophy
at University of Canterbury

by Stephanie Anna Borrie
BSLT (1st Class)
May 2011

DECLARATION BY AUTHOR

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Stephanie Borrie

SUBMITTED WORKS BY THE AUTHOR INCORPORATED INTO THE THESIS

The following referred journal submissions have emanated from the work presented in this thesis:

1. **Borrie, S. A.**, McAuliffe, M. J., & Liss, J. M. (in press). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research.*

   The paper is incorporated into Chapter 1.

2. **Borrie, S. A.**, McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (in press). Familiarization conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes: Special Edition on Speech Recognition in Adverse Conditions.*

   The paper is incorporated into Chapters 3.

PUBLICATIONS RELEVANT TO THE THESIS BUT NOT FORMING PART OF IT

1. McAuliffe, M. J., **Borrie, S. A.**, Good, P. V., & Hughes, L. E. (2010). Consideration of the listener in the assessment and treatment of dysarthria. *ACQuiring Knowledge in Speech, Language, and Hearing*, 12 (1), 16-19.

LIST OF PRESENTATIONS BY THE AUTHOR RELEVANT TO THE THESIS

1. **Borrie, S. A.**, McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2011). Perceptual learning of hypokinetic dysarthria: A follow-up investigation into the mechanisms that underlie passive and explicit learning of dysarthric speech. *Asia Pacific Conference on Speech, Language, and Hearing*, Christchurch, New Zealand: 11-14 January 2011.

2. **Borrie, S. A.**, McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2010). Perceptual learning and intelligibility gains in dysarthria associated with Parkinson's disease. *Health Research Society Canterbury: Seminar Series*, Christchurch, New Zealand: 16 July 2010.

3. **Borrie, S. A.**, McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2010). Learning to better understand distorted speech: An investigation into perceptual learning of dysarthria. *NZSTA Conference: Practice Worth Spreading*, Wellington, New Zealand: 22-23 April 2010.

4. **Borrie, S. A.**, McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2010). A comparison of familiarisation modes on perceptual learning and speech intelligibility of hypokinetic dysarthria. *Fifteenth Biennial Conference on Motor Speech Disorders: Motor Speech Disorders and Speech Motor Control,* Savannah, United States: 4-7 March 2010.

ACKNOWLEDGEMENTS

Hebrews 12:1 (in part) "And let us run with endurance the race God has set before us."

On day one of embarking on my PhD, a colleague informed me that the journey would be like riding a rollercoaster, with its unpredictable ups and downs, twists and turns and the feeling of never quite knowing what was around the next corner. I was told that it was my job to buckle in, grip the sides, and hang on tightly for dear life. This analogy provided little in the way of enthusiasm as I thought about what this may mean for the next three years of my life. I was excited about the PhD journey. I had the opportunity to devote a significant amount of time to research an area about which I was truly passionate. I knew that I would need to work hard, to rise above unanticipated obstacles, and to push beyond perceived limits. But I also knew that it is exactly challenges like this, that bring about some of the most rewarding and memorable moments in life. I know this because I am a runner, and because some of the toughest races that I have ever run, have also been some of the best. So I decided that rather than the rollercoaster ride, I would liken my PhD journey to the goal of running a race, perhaps in this case, it would be a marathon...

The journey begins with my coaches. Without their invaluable input, I would not have made it even a quarter of the way through the intense training that was to come. They helped me to design a plan, a timeline, a way that my goals could be achieved. They were an incredible source of knowledge, inspiration and support. I wish to thank my supervisors Dr Megan McAuliffe and Associate Professor Julie Liss for their absolutely phenomenal coaching over the last three years. Megan and Julie, I cannot thank you enough for the time, energy, and effort that you have invested in me. The insightful discussions, thought-provoking ideas, and perpetual enthusiasm for research have impacted me and my work tremendously. Thank you for encouraging me to aim high. I could not, and would not, have completed this thesis without the both of you. I would also like to extend sincere thanks to Dr Greg O'Beirne for software programming and to Professor Tim Anderson for supporting my participant recruitment endeavours. Your assistance has been absolutely critical to my research.

Successful runners are backed by sponsors, those who are willing to invest their resources to enable the runner's goal to be realised. With this in mind, I would like to acknowledge the many people who willingly gave their time and energy to participate in my research. Without their generosity there simply would have been no data to work with. To all those who participated, I thank you wholeheartedly.

My fellow runners were a tremendous source of encouragement along the way. While we all ran our own race, together we learned to keep running, through rain, hail or shine. To my colleagues—the Creyke Road Gang, the PhD runners, and the Arizona Lab Girls—each one of you has touched my life in a memorable way. Thank you for sharing in my highs and lows and allowing me to share in yours too. You have made the journey a joy to travel and I am truly honoured that our paths have crossed. May our paths continue to cross over and over again.

The marathon is not a fast and hard sprint. The race is long and perseverance is the key. It requires the ability to continually put one foot in front of the next...this would not have been possible without my support crew who stood at the sidelines and cheered loudly for me each and every step of the way. They were undoubtedly the best cheerleaders I could ever have asked for. They held out water when I was thirsty, gave me food when I was weak, and yelled "you can do it" when I started to wonder if I really could. They provided a much needed distraction when it all became too much. My family and friends, I am indebted to your amazing support. Finally, I wish to acknowledge my greatest supporter of all, Glynn. Thank you for your incredible patience, unconditional love, and absolutely unwavering faith in me. Your words of encouragement fuelled me on to the very end.

So I end this PhD feeling extremely thankful. Thankful to all the incredible people who have journeyed with me over the last three years and helped me complete the "race." Above all, I am thankful to God—not only for the race, but for the ability to run.

Stephanie Borrie

ABSTRACT

Perceptual learning, when applied to speech, describes experience-evoked adjustments to the cognitive-perceptual processes required for recognising spoken language. It provides the theoretical basis for improved understanding of a speech signal that is initially difficult to perceive. Reduced intelligibility is a frequent and debilitating symptom of dysarthria, a speech disorder associated with neurological disease or injury. The current thesis investigated perceptual learning of dysarthric speech, by jointly considering intelligibility improvements and associated learning mechanisms for listeners familiarised with the neurologically degraded signal. Moderate hypokinetic dysarthria was employed as the test case in the three phases of this programme of research.

The initial research phase established strong empirical evidence of improved recognition of dysarthric speech following a familiarisation experience. Sixty normal hearing listeners were randomly assigned to one of three groups and familiarised with passage readings under the following conditions: (1) neurologically intact speech (control) (*n* = 20), dysarthric speech (passive familiarisation) (*n* = 20), and (3) dysarthric speech coupled with written information (explicit familiarisation) (*n* = 20). Subsequent phrase transcription analysis revealed that the intelligibility scores of both groups familiarised with dysarthric speech were significantly higher than those of the control group. Furthermore, performance gains were superior, in both size and longevity, when the familiarisation conditions were explicit. A condition discrepancy in segmentation strategies, in which attention towards syllabic stress contrast cues increased following explicit familiarisation but decreased following passive familiarisation, indicated that performance differences were more than simply magnitude of benefit. Thus, it was speculated that the learning that occurred with passive familiarisation may be qualitatively different to that which occurred with explicit familiarisation.

The second phase of the research programme followed up on the initial findings and examined whether the key variable behind the use of particular segmentation strategies was simply the presence or absence of written information during familiarisation. Forty normal

hearing listeners were randomly assigned to one of two groups and were familiarised with experimental phrases under either passive (n = 20) or explicit (n = 20) learning conditions. Subsequent phrase transcription analysis revealed that regardless of condition, all listeners utilised syllabic stress contrast cues to segment speech following familiarisation with phrases that emphasised this prosodic perception cue. Furthermore, the study revealed that, in addition to familiarisation condition, intelligibility gains were dependent on the type of the familiarisation stimuli employed. Taken together, the first two research phases demonstrated that perceptual learning of dysarthric speech is influenced by the information afforded within the familiarisation procedure.

The final research phase examined the role of indexical information in perceptual learning of dysarthric speech. Forty normal hearing listeners were randomly assigned to one of two groups and were familiarised with dysarthric speech via a training task that emphasised either the linguistic (word identification) (n = 20) or indexical (speaker identification) (n = 20) properties of the signal. Intelligibility gains for listeners trained to identify indexical information paralleled those achieved by listeners trained to identify linguistic information. Similarly, underlying error patterns were also comparable between the two training groups. Thus, phase three revealed that both indexical and linguistic features of the dysarthric signal are learnable, and can be used to promote subsequent processing of dysarthric speech.

In summary, this thesis has demonstrated that listeners can learn to better understand neurologically degraded speech. Furthermore, it has offered insight into how the information afforded by the specific familiarisation procedure is differentially leveraged to improve perceptual performance during subsequent encounters with the dysarthric signal. Thus, this programme of research affords preliminary evidence towards the development of a theoretical framework that exploits perceptual learning for the treatment of dysarthria.

KEY WORDS

dysarthria, intelligibility, speech perception, perceptual learning, cognitive-perceptual processes, learning mechanisms, segmental, suprasegmental, linguistic, indexical.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

A Review of the Literature

Borrie, S. A., McAuliffe, M .J., & Liss, J. M. (in press). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research*.

Chapter 1 (section 1.1 and section 1.4) is an adaptation of the manuscript, entitled "Perceptual learning of dysarthric speech: A review of experimental studies," currently in press with Journal of Speech, Language, and Hearing Research. Modifications to the text have been made to ensure consistency and relevance to the current chapter and thesis.

1.1 INTRODUCTION

"The limits of my language stand for the limits of my world."
Ludwig Wittejehstein

Spoken language lies at the very heart of what it is to be human. It is the medium through which one expresses thoughts, feelings and emotions; empowering people to respond to and control their environment (Duffy, 2005). Therefore, it is not surprising that when the ability to communicate effectively is reduced through neurological impairment or disease, profound deleterious effects to an individual's social, family, academic, and vocational life may result (Theodoros, Murdoch, & Goozee, 2001). Communication impairment has been reported as one of the most distressing symptoms of neurological disease (Duffy, 2005).

Dysarthria, a neurological disorder of the motor speech system, manifests itself in perceptual disturbances that compromise the integrity of the speech signal. It commonly results in impaired speech intelligibility. Indeed, intelligibility disturbances have been classified a "hallmark" feature of this speech disorder (Tikofsky & Tikofsky, 1964; Yorkston, Beukelman, & Bell, 1988) and described as "the most clinically and socially important aspects of dysarthria" (Ansel & Kent, 1992, p. 296). As such, treatments that address improving speech intelligibility are fundamental to the successful management of dysarthria.

Speech intelligibility has traditionally been viewed as the property of the speaker (e.g., Black, 1957; Bond & Moore, 1994; Hood & Poole, 1980). Accordingly, dysarthria management has focused primarily upon individual speakers themselves, with emphasis on attempts to improve speech production or equip speakers with strategies or devices to compensate for their impairments (e.g., Duffy, 2005). Recent Cochrane reviews have concluded that there are no high-level studies to support or refute the efficacy of behavioural speech treatment for progressive and non-progressive dysarthrias (Deane, Whurr, Playford, Ben-Shlomo, & Clarke, 2009; Sellars, Hughes, & Langhorne, 2007). Considering the clinical significance of improving intelligibility for individuals with dysarthria, it is critical that research continues to examine the outcomes of behavioural modification on speech production. However, the consideration and development of innovative new forms of treatment is also vital.

Recent definitions of speech intelligibility have highlighted the influence of both speaker and listener in this construct (e.g., Klasner & Yorkston, 2005). With this in mind, Liss (2007) recently proposed a novel remediation strategy for targeting the speech intelligibility impairments exhibited by individuals with dysarthria—specifically, that treatments focus on the listener. The notion of improving a listener's ability to understand the speech of individuals with dysarthria is theoretically based in the broader field of speech perception and more specifically, *perceptual learning*. When applied to speech, perceptual learning describes experience-evoked adjustments to the cognitive-perceptual processes required to recognise spoken language. In brief, these perceptual processes allow the listener to segment a continuous speech stream into individual words (lexical segmentation), to access the lexical items that may match these targets (lexical activation), and to select the most appropriate word for the spoken utterance (lexical competition) (e.g., Jusczyk & Luce, 2002). Put simply, perceptual learning implies that a listener learns to better recognise a speech signal that is initially difficult to understand.

The last decade has seen much research focused on experimental designs that evaluate perceptual learning of speech. There is now a considerable body of evidence regarding the perceptual benefit for listeners familiarised with an ambiguous or unfamiliar speech signal (e.g., time-compressed, noise-vocoded, foreign-accented) (see Samuel & Kraljic, 2009). Research has also begun to investigate this phenomenon with neurologically degraded speech. While the body of research is small, preliminary evidence suggests that perception of dysarthric speech may also improve with training (e.g., D'Innocenzo, Tjaden, & Greenman, 2006; Liss, Spitzer, Caviness, & Adler, 2002). This highlights the potential for perceptual learning to be exploited for rehabilitative gain in dysarthria management. However, if this is to occur, a considerable amount of research is first required. This research must build on existing empirical evidence and develop a theoretical framework for a perceptual learning approach to the treatment of dysarthria. Therefore, this thesis presents a systematic investigation of the theoretical basis for learning to better understand the disordered speech of dysarthria. The focus of the series of studies is hypokinetic dysarthria associated with Parkinson's disease (PD).

The purpose of this introductory chapter is to: (1) describe dysarthria and present a brief overview of the current status of its management; (2) present a short tutorial on models of speech perception and their application to the processing of dysarthric speech; (3) provide a review of perceptual learning[1], both within the broader category of atypical speech and with specific reference to dysarthric populations; and (4) detail the research aims of the current thesis. The introductory chapter concludes with a description of the nature of hypokinetic dysarthria—the dysarthria test case used in the programme of research.

1.2 DYSARTHRIA

Dysarthria refers to impairment in speech motor control arising from neurological disorder or disease (e.g., stroke, traumatic brain injury, and degenerative neurological diseases). In their seminal work, Darley, Aronson, & Brown (1969b) described dysarthria as "...a collective name for a group of speech disorders resulting from disturbances in muscular control over the speech mechanism due to damage of the central or peripheral nervous system" (p. 246). Duffy (2005) extended this definition, stating that dysarthria is characterised by deficits in the speed, strength, range, timing or accuracy of speech movements, with impairment in one or more of the motor subsystems required for speech: respiration, phonation, articulation, resonance, and prosody; and that the resultant weakness, spasticity, incoordination, involuntary movements, or disturbed muscle tone reflect pathophysiologic disturbances. A consistent factor in both definitions is the assumption that the speech symptoms exhibited by individuals with dysarthria are the direct result of the underlying neuropathophysiology—in other words, the perceptual disturbances manifest the classic symptoms of the neurological disease (Weismer & Kim, 2010). Accordingly, disease origins, rather than speech production symptoms, contribute predominately to the differential diagnosis of dysarthria into one of seven major types: spastic, flaccid, ataxic, unilateral upper motor neuron, hyperkinetic, hypokinetic and mixed dysarthria (Darley, Aronson, & Brown, 1969a; Darley, et al., 1969b; Duffy, 2005). Weismer and Kim (2010) recently highlighted that empirical evidence to support the link between disease and dysarthria type is scarce. As such, preference is given to a more recent definition of dysarthria, "a communication deficit, associated with a variety of neurological diseases, in which speech movements are affected in

[1] Perceptual learning is reviewed with respect to experimental studies that have examined manipulation of the listener experience (familiarisation/training).

such a way as to make the speech acoustic signal unsuitable, in varying ways and degrees, for language perception" (Weismer, 2006, p. 320).

1.2.1 Prevalence and Incidence of Dysarthria

Dysarthria is common in neurological impairment and disease. Identified as the primary diagnosis of 54% of 10,444 patients evaluated in the Section of Speech Pathology in the Department of Neurology at the Mayo Clinic from 1987-1990 and 1993-2001, dysarthria represents one of the most prevalent acquired neurologic communication disorders (Duffy, 2005). While data reporting the prevalence and incidence of dysarthria in the general population is limited, insight into the pervasiveness of this motor speech disorder is evident in a number of studies. It is estimated that approximately 20 to 30% of individuals post-stroke (Arboix, Marti-Vilalta, & Garcia, 1990; Melo, Bogousslavsky, van Melle, & Regli, 1992; Warlow et al., 1996), and between 10 and 65% of individuals with traumatic brain injury (Sarno, Buonaguro, & Levita, 1986; Yorkston, Honsinger, Mitsuda, & Hammen, 1989) will exhibit dysarthria. With the progression of degenerative disease, approximately 60 to 80% of those with PD (Adams, 1997) and 50% of individuals with multiple sclerosis (Harteliusa, Runmarkerb, & Andersenb, 2000; Sandyk, 1995) will develop dysarthria in some form. In the case of Amyotrophic lateral sclerosis, it is reported that dysarthria will affect the majority of individuals as the disease progresses (Wijesekera & Leigh, 2009).

1.2.2 Speech Characteristics of Dysarthria

Any disturbance in the sensorimotor processes that underlie speech production may manifest perceptually in the speech signal. The resultant perceptual deficits that characterise dysarthric speech are commonly referred to as *deviant perceptual features* (Duffy, 2005). These features may be classified according to the specific speech subsystem (e.g., respiration or phonation) and/or with regards to the segmental and suprasegmental properties of speech they affect. For example, when classified via speech subsystem deficits, a breakdown at the respiratory level of speech production may be evident in audible inspiration, inhalatory strider and/or grunting during speech production; phonatory system impairments may be reflected in a harsh, hoarse, breathy or strained-strangled sounding voice; articulatory disturbances may manifest in imprecise phoneme production; disturbed prosody may be seen in aberrant stress patterns and speaking rate; and resonatory deficits can be observed in the presence of insufficient or excess nasality during speech (Duffy, 2005).

More recently, the deviant perceptual features of dysarthria have been classified according to their segmental and suprasegmental features (Liss, 2007). Segmental features refer to individual phoneme and syllable productions within a spoken word. Accordingly, segmental errors include phoneme omissions, distortions and substitutions, as well as co-articulatory disturbances across phoneme strings (Liss, 2007). In comparison, suprasegmental properties reflect the more global aspects of speech production and describe vocal components that extend over more than one sound segment in an utterance. Suprasegmental errors are evident in the parameters of intonation, vocal intensity, and rate-rhythm. The deviant perceptual features that are associated with dysarthria can, individually or in combination, significantly impact upon intelligibility of the speech signal.

1.2.2.1 Intelligibility Impairments

The term intelligibility refers to how effectively the speech signal can be understood by a listener. In the field of dysarthria, it is used as an index of severity of the speech disorder or to describe the functional limitation afforded by the speech impairment (e.g., Hustad, 1996; Yorkston, Beukelman, Strand, & Bell, 1999). Measures of speech intelligibility are also frequently used to document treatment effects (e.g., Kennedy, Strand, & Yorkston, 1994; Yorkston, Hammen, Beukelman, & Traynor, 1990). Regardless of dysarthria sub-type, it is anticipated that the majority of speakers with dysarthria will experience reduced intelligibility to some degree (Darley, et al., 1969b; Yorkston, et al., 1999). With adverse effects on the success, competence, and effectiveness of communication, the intelligibility impairments that characterise dysarthria can significantly impact upon quality of life (e.g., Bunton & Weismer, 2001; Hustad, Beukelman, & Yorkston, 1998). As such, improvement to speech intelligibility is considered a fundamental goal of dysarthria rehabilitation and management (Ansel & Kent, 1992; Yorkston, et al., 1999).

1.2.3 Dysarthria Management

There is no one approach to addressing the needs of the individual with dysarthria. Rather, management is generally targeted through multiple modalities and is highly dependent upon the needs and presenting speech characteristics of the individual. While behavioural management is considered central to dysarthria rehabilitation, additional approaches may include medical interventions, prosthetic management, and alternative and augmentative communication (AAC) (Duffy, 2005). Medical interventions consist of pharmacological and surgical treatments, which may have a direct (e.g., botulinum toxin to treat spasmodic dysphonia) or indirect (e.g., dopaminergic medications for PD) influence upon speech production (Duffy, 2005). Improvements in speech may also be achieved through the use of prosthetic (e.g., palatal lift prosthesis) or assistive devices (e.g., voice amplifiers). When speech is most severely compromised, management may involve devices (e.g., alphabet charts, electronic talking devices) and/or strategies (e.g., gesture) to augment or substitute for speech.

1.2.3.1 Behavioural Management

The overarching goal of the behavioural management of dysarthria is to maximise communication (Duffy, 2005). Accordingly, such interventions are numerous and wide ranging. Behavioural interventions are divided into *speaker-oriented* and *communication-oriented* approaches. The speaker-oriented approach has traditionally played a predominant role in the management of dysarthria. Treatment is focused on the individual speaker and his/her speech signal. It aims to improve speech production or equip a speaker with strategies and/or devices to compensate for their deficits (Duffy, 2005; Rosenbek & LaPointe, 1985; Yorkston, et al., 1999). Speaker-orientated treatment may focus on reducing, or compensating for, underlying motor deficits and can involve both speech and non-speech activities (Duffy, 2007). Communication-oriented approaches, by comparison, involve efforts independent of the speech signal. These include modifications to the communication environment, supplementation techniques (including the use of AAC), and strategies to assist communication interaction. A comprehensive review of the numerous behavioural approaches which have been employed in dysarthria rehabilitation is provided in Duffy (2005).

1.2.3.2 Treatment Efficacy

While research evidence to support the use of speaker- and communication-orientated approaches for rehabilitation in dysarthria is growing, currently there is only limited high level efficacy data for such treatments. Indeed, recent Cochrane reviews have concluded that there are no high level studies to support or refute the efficacy of speech treatment for progressive and non-progressive dysarthrias (Deane, et al., 2009; Sellars, et al., 2007). Systematic reviews in the areas of respiratory-phonatory dysfunction (Spencer, Yorkston, & Duffy, 2003), velopharyngeal function (Yorkston et al., 2001), spasmodic dysphonia (Duffy & Yorkston, 2003), and speech supplementation techniques (Hanson, Yorkston, & Beukelman, 2004) have reached similar conclusions, again stating that the evidence base for interventions that address these aspects of dysarthria rehabilitation is limited.

A small number of published reports do provide high-level scientific evidence of positive treatment outcomes for individuals with dysarthria. However, further research into a number of aspects including the application of such approaches across dysarthria subtypes, long-term maintenance of treatment effects, and generalisation across settings is required (Hustad & Weismer, 2007; Spencer, et al., 2003; Yorkston, Hakel, Beukelman, & Fager, 2007; Yorkston, et al., 2001). Other behavioural approaches have good face value but are supported by anecdotal endorsements only (Duffy, 2005). Although, it appears that clinically, behavioural intervention provides benefits for patients and their families, a strong empirical basis for these approaches is still to be established.

Taken together, the pervasiveness of dysarthria, the debilitating associated intelligibility impairment, and the limitations of the current treatment efficacy data highlight the clear need for ongoing research in the field of dysarthria management. While it is crucial that research continues with efforts to document the outcomes of existing behavioural approaches, it is also important that research investigates new avenues and novel approaches to the treatment of dysarthria. As improved intelligibility is considered one of the primary goals of dysarthria management, the foundation of any innovative approach should begin with this in mind.

1.2.3.3 Future Directions in Dysarthria Management

The word is half his that speaks, and half his that hears it.
Michel de Montaigne

Speech intelligibility is defined as “the accuracy with which a message is conveyed by a speaker and recovered by a listener” (Klasner & Yorkston, 2005, p. 127). This definition underscores the essential role of both speaker and listener in the communication process. With the speaker-listener process in mind, it is proposed that the intelligibility impairments exhibited by individuals with dysarthria may benefit from treatments that focus on the listener (Liss, 2007). While conceptually, listener-targeted remediation in dysarthria is novel, its potential should not be underestimated. Dysarthria very rarely occurs in isolation. Physical, cognitive and memory deficits frequently co-occur, all of which can greatly reduce the individual’s capacity to learn and maintain benefits from speaker-oriented interventions (Duffy, 2005). However, treatment that focuses on the neurologically intact listener (e.g., family members, friends, carers), thereby bypassing the speaker and any associated conditions that may adversely affect treatment gains, may prove key to optimising communication success in those with dysarthria (McAuliffe, Borrie, Good, & Hughes, 2010). The potential for listener-focused treatments, whereby the listener may be trained to better understand dysarthric speech, is theoretically based in the fields of speech perception and perceptual learning.

1.3 SPEECH PERCEPTION

Speech perception has been defined as “the process of imposing a meaningful perceptual experience on an otherwise meaningless speech input” (Massaro, 2001, p. 14870). It is a broad term which encompasses a number of perceptual processes that facilitate comprehension and interpretation of individual spoken words, or a collection of words within a spoken phrase or sentences (i.e., connected speech). Knowledge of the processes that underpin a listener’s ability to perceive spoken language is critical for an understanding of perceptual learning of speech. Accordingly, this section begins with a review of the basic processes of speech perception.

1.3.1 Spoken Word Recognition

On the surface, spoken word recognition appears a seemingly effortless task. However, the identification of a spoken target must be selected from literally hundreds of thousands of possibilities. Over the past four decades, significant research efforts have been devoted to detailing the processes that enable a listener to carry out the task of spoken word recognition. The development of theoretical models to account for recognition of spoken words reveals that it is in fact a highly complex skill. Some of the more influential models of spoken language recognition include the Logogen model (Morton, 1969), the Cohort theory (Marslen-Wilson & Welsh, 1978), the TRACE model (McClelland & Elman, 1986), the Shortlist model (Norris, 1994), the Neighbourhood Activation model (Luce & Pisoni, 1998), and the PARSYN model (Luce, Goldinger, Auer, & Vitevitch, 2000). A detailed account of each of these models is beyond the scope of this thesis; instead, the reader is directed to Liss (2007) for a summary of each of the models identified above. While varying in their specifics, all of the models commonly assume that two fundamental perceptual processes underlie the recognition of spoken words: lexical activation and lexical competition.

1.3.1.1 Lexical Activation

Lexical activation is the initial process involved in spoken word recognition. As a word is produced, a cohort of lexical items with similar acoustic features to that of the spoken word is activated (e.g., Connine, Blasko, & Wang, 1994; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson, 1990; Stevens, 2004; Zwitserlood, 1989). Processes of lexical activation begin as soon as the initial phonemes of the spoken word are produced, and as such, the role of word onsets in activation have received considerable attention (e.g., Marslen-Wilson & Welsh, 1978; Nooteboom, 1981). For example, upon presentation of the syllable onset *el*, items such *elephant, elevator* and *elementary* may be activated and identified as possible lexical candidates for later word recognition (P. Warren & Marslen-Wilson, 1987, 1988). Early descriptions of lexical activation postulated that activated items included those with similar word onsets (Marslen-Wilson & Welsh, 1978) and later extended to also include items with similar stressed syllable patterns (Cutler & Norris, 1988). However, most recent models of lexical activation contend that any consistency between input and representation can facilitate activation of lexical items. For example, word-final position can contribute to the activation of lexical items, which is particularly relevant where the target word attains its uniqueness during the final segment of its production (e.g., *batter-batten*) (Connine, Blasko, & Titone, 1993; Connine, et al., 1997). Other lexical activation cues may arise in similarities in rhyme (e.g., *speaker-beaker*) (Allopena, Magnuson, & Tanenhaus, 1998) and phonetic properties, even when there is no shared position-specific segment (e.g., *shun-gong*) (Luce, et al., 2000).

1.3.1.2 Lexical Competition

Once a cohort of lexical possibilities is activated, the process of lexical competition may then proceed. Word items that are viewed as more likely candidates are facilitated, and items that are regarded as less likely candidates are discarded (Liss, 2007). This process of facilitation and inhibition continues until a final word remains. The remaining word is selected and recognition of the word occurs. Thus lexical competition results in the identification of the lexical candidate that best matches the spoken input.

The efficiency and accuracy with which this process transpires is influenced by lexical neighbourhood density (the number of words which are similar to the target) (Luce, et al., 2000; Luce & Pisoni, 1998) and word frequency effects (how frequently the target occurs in everyday speech) (Luce, et al., 2000). Sub-lexical effects such as listener expectation based on co-articulatory acoustic cues (Davis, Marslen-Wilson, & Gaskell, 2002; P. Warren & Marslen-Wilson, 1987, 1988) and positional probabilities of phonemes (Luce, et al., 2000; Luce & Large, 2001) also have the potential to affect the competition process.

1.3.2 Recognition of Connected Speech

While the processes of activation and competition among lexical items continue to operate during the recognition of connected speech (i.e., spoken phrases or sentences), these processes are subject to further influence by the intended meaning of the message (e.g., Samuel, 1981). With connected speech, the semantic and syntactic relationships between spoken words cultivate a source of lexical predictability. This enables the listener to tolerate a degree of acoustic ambiguity not afforded by the recognition of words produced in isolation. A well-known phenomenon, phonemic restoration, highlights this very point. In their seminal work, Warren and Warren (1970) showed that when specific phonemes had been extracted from a short passage and replaced with a cough sound, listeners claimed to have heard the phonemes presented, that is, the absent phonemes and the cough production—in sum, listeners restored the absent phonemes. These studies, along with others (Garnes & Bond, 1976), provided convincing evidence that the perception of words within an utterance is strongly influenced by the lexical predictability provided within the spoken message. However, in order to parse connected speech into individual word items, an additional process, *lexical segmentation,* is required. Figure 1.1 contains a pictorial representation of spoken word recognition and the interplay of the perceptual processes: lexical segmentation, lexical activation, and lexical competition.

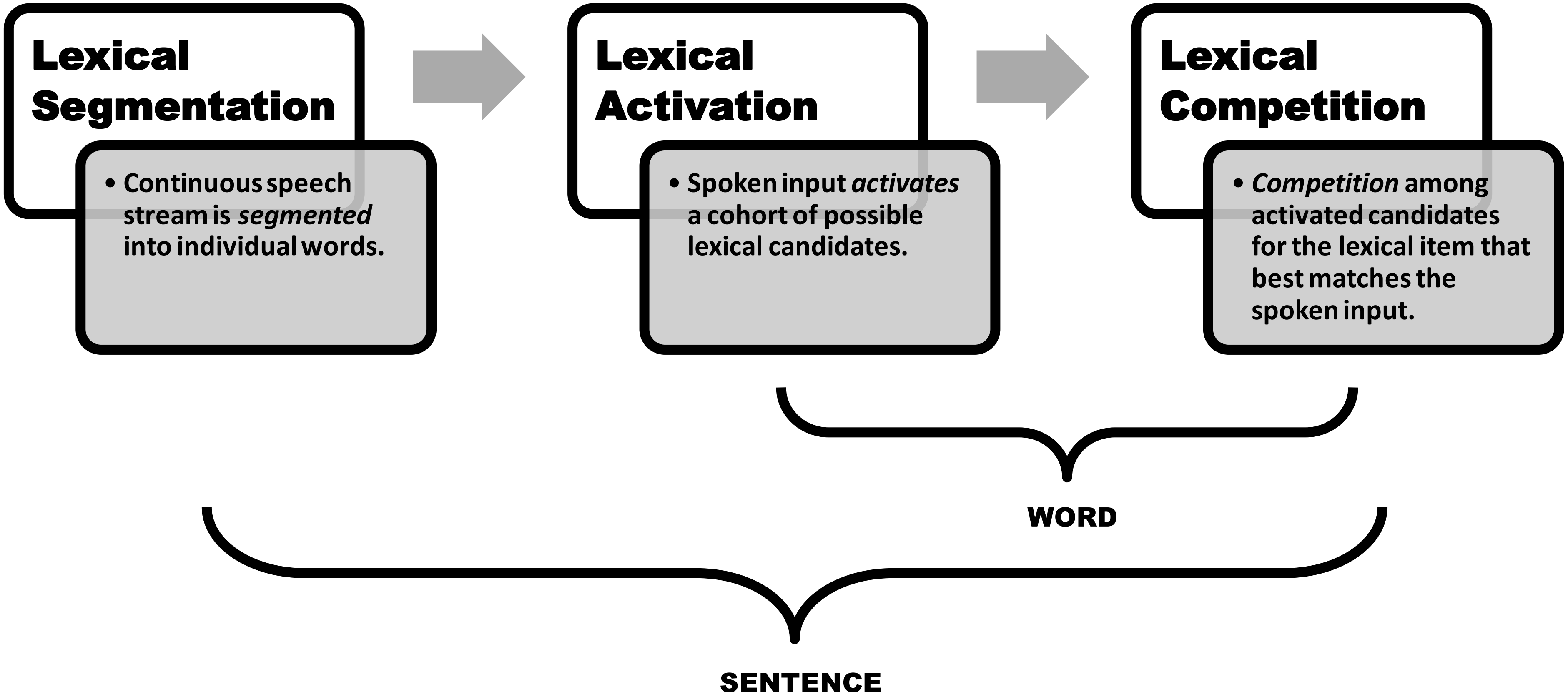

Figure 1.1. Perceptual processes involved in the recognition of spoken language.

1.3.2.1 Lexical Segmentation

Connected speech is comprised of a continuous stream of acoustic energy. From this continuous acoustic stream, a listener must determine where one word ends and where the next begins. This process of identifying word boundaries is achieved by lexical segmentation, a language-specific perceptual process considered central to the understanding of spoken language recognition (e.g., Cutler, Mehler, Norris, & Segui, 1986). Early literature examining the process of speech segmentation reported two distinct and contrasting theories to account for its occurrence; *lexically driven* processes and *sublexically driven* processes. Lexically driven processes are dependent on knowledge of the structure of language (e.g., semantic plausibility and syntactic rules) and lexical items are activated in line with the rules and regularities of language (e.g., McClelland & Elman, 1986; Norris, 1994; Stevens, 2004). In contrast, sublexically driven processes rely on sublexical cues including prosody and syllabic stress (Cutler & Norris, 1988), phonotactic probabilities (e.g., McQueen & Cutler, 1998), allophonic variations (e.g., Nakatani & Dukes, 1977), and expectations concerning the nature of coarticulatory productions (e.g., Davis, et al., 2002) to identify the beginning and ends of a word. For example, Culter and colleagues (1992; 1988) proposed the Metrical Segmentation Strategy (MSS), wherein listeners will exploit prosodic properties of the signal to predict the location of word-onsets in connected speech when phonemic ambiguity is high (see also Mattys, White, & Melhorn, 2005). Based on the statistical probabilities of English, it is thought that speech segmentation will be largely successful if listeners treat strong syllables (those receiving relative stress through longer duration, fundamental frequency change, increased loudness, and relatively full vowel) as word onsets in a stream of connected speech (Cutler & Carter, 1987). Evidence of this strategy can be found in listener's lexical boundary error (LBE) patterns, manifested in the tendency to mistakenly insert lexical boundaries before strong syllables and to delete boundaries before weak syllables.

Most recently, it has been proposed that lexical segmentation involves the integration of both lexical and sublexical processing strategies (McQueen & Cutler, 2001b). Framed within this integrative model, Mattys and colleagues (2005) have suggested that the use of lexical and sublexical information to segment speech may depend on the quality of acoustic signal and the richness of contextual information to which the listener is exposed. This hierarchical model of lexical segmentation postulates that listeners will employ low-cost lexically driven strategies when the speech signal quality is high and/or context is rich and will integrate the higher-cost sublexically driven strategies as required. Specifically, in instances where segmental information is insufficient to afford any cues for lexical segmentation, listeners will utilise suprasegmental properties to inform word boundary decisions (Mattys, et al., 2005).

Speech perception research has traditionally focused on the *linguistic* aspects of the speech signal (e.g., Luce, et al., 2000; McClelland & Elman, 1986; Norris, 1994). Linguistic properties refer to the phonological, morphological, syntactic, and semantic elements expressed through the word, phrase, and sentences of a spoken message (Levi & Pisoni, 2007). However, the speech signal carries *indexical* information as well. This information occurs independent of the spoken message and provides details pertaining to specific attributes of the speaker (e.g., Hagiwara, 1997; Van Lancker, Kreiman, & Emmorey, 1985). The influence of indexical information in processing of spoken language is now being acknowledged.

1.3.3 Recognition of Indexical Information

Indexical properties of speech refer to speaker-specific attributes that afford information regarding the individual's identity (e.g., Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985), gender (e.g., Munson, McDonald, DeBoe, & White, 2006), regional dialect (e.g., Hagiwara, 1997; Hillenbrand, Getty, Clark, & Wheeler, 1995), and emotional state (e.g., Costanzo, Markel, & Costanzo, 1989; Frick, 1985; Murry & Arnott, 1993; Scherer, Banse, Wallbott, & Goldbeck, 1991). They manifest acoustically in features such as fundamental frequency, formant spacing, breathiness, relative segment durations, overall speaking rate, and vocal effort (Nygaard, 2008). Indexical signal properties introduce substantial variability, both within and between speakers, and can profoundly influence the acoustic realisations of speech (e.g., Hillenbrand, et al., 1995; Nearey, 1978; Nygaard, 2008; Peterson & Barney, 1952). Traditional accounts of speech perception have focused predominantly upon on the processing of linguistic information, largely ignoring the potential contributions of indexical properties of the signal to speech perception (e.g., Luce, et al., 2000; Luce & Pisoni, 1998; Morton, 1969; Norris, 1994). While acknowledging the existence of indexical properties, traditional theoretical paradigms contend that such information is processed independently of linguistic information (e.g., Halle, 1985). It has been assumed that the perceptual system will recruit a *normalisation* process to enable it to deal with the enormous acoustic variation imposed by individual speaker productions (e.g., Brown & Carr, 1993; Halle, 1985; Joos, 1948). Under normalisation, the perceptual system removes any distinctive and variable features imposed by the speaker, reducing the acoustic signal down to its most pure and normalised linguistic form. With an abstractly defined, stable representation of the linguistic information imprinted within a listener's memory, speech perception can continue successfully in the face of substantial individual acoustic variability (see Goldinger, 1998; Tenpenny, 1995 for reviews).

However, these conventional models have been challenged by research which demonstrates that, rather than being discarded in the process of perception, indexical properties may play a key role in the recognition of spoken language (e.g., Mullennix & Pisoni, 1990; Summerfield, Haggard, Foster, & Gray, 1984). Evidence that linguistic processing is influenced by indexical properties is demonstrated in a number of studies that observe a perceptual benefit with indexical consistency. Creelman (1957) correlated word recognition in noise with percent recognition accuracy when word lists were produced under single- versus multiple-speaker conditions. They found an inverse relationship between intelligibility and number of speakers: word recognition increased as speaker numbers decreased. Mullennix, Pisoni and Martin (1989) replicated the findings and observed increased speed and accuracy of word recognition under single- verses multi-speaker conditions. A number of other studies have also demonstrated the perceptual advantage of single-speaker conditions (e.g., Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Palmeri, Goldinger, & Pisoni, 1993) and findings extend to hearing-impaired adults (Kirk, Pisoni, & Miyamto, 1997) and preschool children (Ryalls & Pisoni, 1997). Evidently, perception of spoken language is influenced by indexical variation. This has led to the proposal that speech perception is in fact a highly integrated process, whereby indexical properties of lexical items are retained and encoded alongside linguistic information (Goldinger, 1996, 1998; Pisoni, 1997). The development of theoretical models of speech perception must, therefore, account for the role of acoustic variability, both linguistic and indexical, evident in spoken language (Nygaard, 2008).

Degraded speech represents an extreme form of acoustic variation. With dysarthria, this variability can manifest in both linguistic (e.g., irregular articulatory breakdowns) and indexical (e.g., variable speech rate) properties of the signal (see section 1.2.2). To date, an account of how the perceptual system deals with the neurologically degraded speech signal is yet to emerge.

1.3.4 Recognition of Dysarthric Speech

While significant research efforts have been devoted to the study of spoken language recognition, few studies have attempted to apply models of speech perception to recognition of dysarthric speech. This is somewhat surprising, given the recognised role of the listener in speech intelligibility (e.g., Hustad & Beukelman, 2001; Yorkston, Strand, & Kennedy, 2006). Recently however, researchers have begun to direct their attention towards theoretical models of speech perception, and the insights these models may offer to the perception of dysarthric speech. It has been postulated that the acoustic degradations that characterise dysarthric speech are likely to interfere with the fundamental processes involved in speech perception: lexical activation, lexical competition, and lexical segmentation (Liss, 2007). For example, phonemic distortions may activate larger than necessary lexical cohorts and increase cognitive demands of lexical competition (Liss, 2007; Nusbaum & Schwab, 1986). Furthermore, segmental ambiguity may necessitate reliance on sublexical cues to achieve lexical segmentation, yet the deviant speech features observed in dysarthric speech may hinder the success of applying such strategies.

To date, perhaps the most pertinent research regarding the application of models of spoken language processing to the perception of dysarthric speech is found in two successive studies by Liss and colleagues (1998; 2000a). These studies offer insight into the perceptual challenge that different types of dysarthric speech pose to the process of lexical segmentation. In the first study, a group of 70 listeners transcribed experimental phrases produced by speakers with hypokinetic dysarthria associated with Parkinson’s disease (Liss, et al., 1998). Hypokinetic dysarthria is characterised by a rapid speaking rate, monotony, monoloudness, and phoneme imprecision (Darley, et al., 1969b), all of which serve to diminish syllable strength contrastivity. The listener’s orthographic transcriptions were analysed for percent words correct (PWC) to obtain an index of the overall intelligibility of the dysarthric signal. Analysis of LBE patterns was also conducted on the transcripts. Based on the MSS predictions (see section 1.3.2.1), LBE analysis enables segmentation errors to be interpreted relative to syllable stress. Findings of this suprasegmental level error pattern analysis revealed that listeners made a large number of lexical segmentation errors and the pattern of errors was generally consistent with the predictions offered by the MSS, in which boundary decisions were guided by prosodic cues. In addition, when individual speaker data was analysed, the study found that listeners were less successful at applying the predicted segmentation strategies with speakers exhibiting the greatest acoustic evidence of decreased syllabic contrastivity. These findings provided support for the initial predictions that the reduced syllabic stress cues that are typically evident in speakers who exhibit hypokinetic dysarthria would interfere with processes of lexical segmentation.

A follow-up study compared listener processing of two different types of dysarthric speech that differed significantly in their perceptual presentation: hypokinetic dysarthria (rapid rate, monotony, reduced vowel working space) and ataxic dysarthria (tendency toward syllabic isochrony, excessive loudness variation, and reduced vowel working space consequent to reductions in vowel strength) (Liss, et al., 2000a). A group of 60 listeners transcribed 60 experimental phrases produced by speakers of a particular speech type (hypokinetic, ataxic, or control), with n = 20 in each group. Analysis of the transcripts was identical to the earlier study (Liss, et al., 1998). Examination of the LBEs made by listeners transcribing dysarthric speech revealed different error patterns for lexical segmentation of hypokinetic versus ataxic speech. While listeners transcribing hypokinetic speech conformed to MSS predicted error patterns, the patterns for listeners transcribing ataxic speech revealed a weak or lack of adherence to the expected patterns. The authors hypothesised that the two types of dysarthria posed different perceptual challenges for the listener. Specifically, it was proposed that the prosodic deficits that characterise the ataxic speech signal may interfere with the processes of lexical segmentation to a greater degree than the prosodic deficits that characterise the hypokinetic speech signal. These studies offer insight into some of the challenges that may arise with employing typical perceptual processes to understand neurologically degraded speech.

Recent work by Mattys and Liss (2008) took this one step further and investigated the effect of indexical variability on the perception of hyperkinetic dysarthria. In this study, 72 listeners were assigned to one of three speech types (control, mild dysarthria, severe dysarthria), with n = 24 in each group. Listeners were presented with two consecutive blocks of speech stimuli: (1) 60 words (half produced by a female speaker and the other half by a male speaker), (2) 40 of the same words from block one (half produced in the same voice and the other half in the other voice) and 20 different words (half produced by a female speaker and the other half by a male speaker). Listeners were required to decide if each word in the second block had been played in the initial block ("old" or "new"), regardless of whether the word was produced by the same voice. The authors observed that, for all three speech types, words were better recalled if played in the same voice, as opposed to a different voice between the two successive blocks. In addition, the perceptual advantage of indexical consistency was significantly greater for listeners recalling words produced by speakers with dysarthria relative to neurologically intact controls. The findings extend support for models of spoken word recognition in which indexical information informs linguistic processing (e.g., Palmeri, et al., 1993). Furthermore, it appears that speaker-specific detail may be particularly important when listeners are challenged by the speech of individuals with dysarthria.

These three studies provide the initial foundations for the development of a theoretical account that describes perceptual processes involved in the recognition of dysarthric speech. There is preliminary evidence to suggest that while the dysarthric signal may interfere with the fundamental processes of speech perception, models of typical speech recognition may offer some application to the processing of neurologically degraded speech (e.g., Mattys & Liss, 2008). However, any model that attempts to account for the processing of a degraded or atypical speech signal must also appreciate the adaptive nature of the perceptual system. Accordingly, a discussion of perceptual learning ensues.

1.4 PERCEPTUAL LEARNING

Defined as "relatively long-lasting changes to an organism's perceptual system that improves its ability to respond to its environment and are caused by this environment" (Goldstone, 1998, p. 585), perceptual learning of speech refers to the experience-evoked capacity to adapt or retune the speech perception system. That is, when listeners are familiarised with a speech signal that is unfamiliar or ambiguous, they are able to modify their perceptual strategies for subsequent processing of the atypical speech (Samuel & Kraljic, 2009). Based on interactive models of speech perception, it is proposed that an individual's perceptual system is flexible, and dynamically adjusts to match the information provided in the incoming signal (e.g., McClelland & Elman, 1986).

Evidence for an adaptable speech perception system has been demonstrated in numerous studies that have reported perceptual learning for listeners familiarised with atypical speech (e.g., Bradlow & Bent, 2008; Davis, Johnsrude, Herrvais-Adelman, Taylor, & McGettigan, 2005; Eisner & McQueen, 2005; Francis & Nusbaum, 2009; Francis, Nusbaum, & Fenn, 2007; Golomb, Peelle, & Wingfield, 2007; Greenspan, Nusbaum, & Pisoni, 1988; Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003). Based on these findings, it has been speculated that perceptual learning may be one avenue by which listeners could learn to better process dysarthric speech and, hence ultimately, result in the development of a treatment that facilitates improved outcomes for individuals with dysarthria (Liss, 2007). However, few studies have investigated perceptual learning of the neurologically degraded speech signal. Furthermore, those that have, yield equivocal findings (e.g., Garcia & Cannito, 1996; Liss, et al., 2002). Accordingly, an overview of perceptual learning with atypical speech is provided before reviewing its application to dysarthric speech.

1.4.1 Perceptual Learning of Atypical Speech

The laboratory study of perceptual learning has revealed important information about the ways in which familiarisation with atypical speech alters perception. At the phoneme level, it has been shown that perceptual shifts in phoneme category boundaries occur following experience with ambiguous tokens embedded within lexical contexts (e.g., Eisner & McQueen, 2005; Eisner & McQueen, 2006; Kraljic & Samuel, 2005, 2006; Norris, et al., 2003). For example, Norris et al. (2003) observed that when Dutch listeners were presented with an ambiguous phoneme (acoustically and perceptually halfway between /s/ and /f/) in real or non-word contexts, listeners were able to extend the boundaries of one of their internal fricative categories (/s/ or /f/) to include the ambiguous phoneme. That is, listeners' internal representations of the acoustic information constituting of /s/ or /f/ shifted to accommodate the ambiguous phoneme. The nature of the learning attributed to the phenomenon of category shifting has been termed perceptual adaptation, whereby training facilitates an acoustic-phonetic re-mapping of phonological information at the segmental level of perceptual processing (e.g., Eisner & McQueen, 2005).

Perceptual learning effects have also been reported as improvements in intelligibility (word recognition accuracy) of atypical speech following a familiarisation experience. These unfamiliar or degraded acoustic signals can vary significantly along multiple phonetic and/or prosodic dimensions to that of typically encountered speech. Intelligibility benefits have been demonstrated in listeners who received training with foreign-accented (e.g., Bradlow & Bent, 2008; Weill, 2001) and hearing-impaired speech (e.g., McGarr, 1983), as well as artificially manipulated acoustic signals such as noise-vocoded (e.g., Davis & Johnsrude, 2007; Davis, et al., 2005), computer-synthesised (Francis & Nusbaum, 2009; Greenspan, et al., 1988; Hoover, Reichle, Van Tasell, & Cole, 1987), and time-compressed (e.g., Golomb, et al., 2007) speech. As with phonemic category shift research, it is postulated that the source of perceptual benefit occurs primarily at the segmental level of perceptual processing. When listeners are exposed to the atypical speech pattern, the unique and systematic acoustic-phonetic characteristics of the atypical signal are mapped onto a listener's existing phonological space, causing a shift in perceptual representation of particular phonemes (e.g., Dupoux & Green, 1997; Francis, et al., 2007). This shift is thought to benefit the cognitive-perceptual processes of speech perception, particularly lexical activation (e.g., reduced activation of a larger than necessary lexical cohort) and lexical competition (e.g., reduced competition for processing resources and increased likelihood of correct target selection), thereby yielding improved intelligibility.

Based on a number of findings, the most plausible account for these segmental benefits is that familiarisation with the atypical signal induces an attentional shift toward more phonetically informative acoustic cues (e.g., Francis & Nusbaum, 2000; Francis, et al., 2007; Nusbaum & Goodman, 1994; Pisoni, Lively, & Logan, 1994). According to this explanation, familiarisation with the atypical signal does not increase the quality or the quantity of the available acoustic information, but rather directs cognitive resources to those cues considered most relevant for recognition of the unique speech. For example, Francis et al., (2000) provided empirical evidence that the provision of category-level feedback for listeners familiarised with synthetic speech provoked changes in the way in which place of articulation cues were exploited. More recently, Francis and Nusbaum (2009) observed a relationship between working memory and perceptual learning, wherein listeners familiarised with synthetic speech were better able to utilise working memory for improved recognition of the atypical signal. If a familiarisation experience does in fact improve the distribution of attentional resources (i.e., increased attention toward more informative cues at the expense of less relevant information), demands on working memory may decline, and improved recognition may result (Francis, et al., 2007).

Perceptual learning research using time-compressed speech, a signal characterised by systematic manipulation to its temporal characteristics, has demonstrated that listeners may also learn something about the global prosodic features of the speech signal—specifically its rhythmic qualities (e.g., Pallier, Sebastian-Galles, Dupoux, & Christophe, 1998; Sebastian-Galles, Dupoux, Costa, & Mehler, 2000). The mechanism for this learning may be described as *rhythmic expectancy*, whereby listeners can anticipate and focus attention on high-yield aspects of the signal when they have adapted to the systematically varied rate and rhythm. Sebastian-Galles and colleagues (2000) examined perceptual learning of time-compressed speech across different language classes with distinguishably different rhythmic patterns (syllable-timed vs. stress-timed vs. mora-timed). They found that perceptual learning outcomes were influenced by the rhythmic properties of the training signal. For example, familiarisation with syllable-timed languages facilitated improved processing of other syllable-timed languages, but not with signals that differed in rhythmic patterns. This suggests that acoustic-phonetic remapping is not the only source of benefit that underlies experience-evoked intelligibility improvements and that suprasegmental learning may facilitate subsequent lexical segmentation of speech with similar rhythmic structure.

Traditionally assumed to have limited relevance to linguistic processing (e.g., Halle, 1985), a role for indexical information in perceptual learning of speech has recently been acknowledged (e.g., Loebach, Bent, & Pisoni, 2008). Nygaard and colleagues (1994) found that listeners trained to identify the names of ten unfamiliar speakers exhibited significantly greater recognition scores when presented with novel words produced by these same speakers, relative to listeners presented with novel words produced by unfamiliar speakers. Similar perceptual benefits afforded by attention to indexical properties of the signal were observed with sentence-level recognition in a follow-up study (Nygaard & Pisoni, 1998). In addition, the benefit of speaker familiarity on subsequent linguistic processing has been replicated with older individuals (Yohan & Sommers, 2000). More recently, Loebach et al., (2008) revealed that the perceptual benefit of training on indexical properties may also extend to the perception of noise-vocoded speech. Listeners engaged in a speaker identification task made significant intelligibility improvements and furthermore, the performance gains were as great as those achieved by listeners engaged in a linguistic-based transcription training task. Thus, these studies generate preliminary evidence that indexical information may also inform recognition of artificially degraded speech.

Taken together, it appears likely that multiple potential sources of perceptual learning exist. While the evidence regarding learning sources and the relative contribution of different levels of information is limited, it may be presumed that familiarisation with atypical speech enables listeners to extract something about the unusual regularities of the signal, and that this facilitates improved perceptual processing in subsequent encounters. Until now, this tutorial has treated *familiarisation* or *training* with atypical speech in a rather nebulous way. However, the specific ways in which listeners receive training vary on a number of levels including familiarisation material, familiarisation conditions, and amount of familiarisation. Such factors may or may not influence the longevity of learning and whether effects are generalised across stimuli and/or speakers. These characteristics of perceptual learning are discussed in turn.

1.4.1.1 Familiarisation Material

Familiarisation material describes the stimuli (usually speech) used to promote perceptual learning of the speech signal. Studies have reported that perceptual learning may be most robust when listeners are familiarised with real word, rather that nonword, stimuli (e.g., Davis, et al., 2005; McQueen & Mitterer, 2005; Norris, et al., 2003). This suggests a lexical influence in perceptual learning of speech. When listeners were familiarised with an ambiguous phoneme embedded within word or nonword training material, category boundary shifts were identified only for those listeners trained with real words (Norris, et al., 2003). Using noise-vocoded speech, a signal characterised by systematic manipulation to its spectral information, similar findings regarding the benefit of lexical information were reported (Davis, et al., 2005). Listeners trained with sentences containing real words demonstrated improved word recognition of the noise-vocoded speech signal, whereas the learning response was not identified for listeners familiarised with nonword sentence stimuli. When the familiarisation material was further manipulated to remove sentence-level or syntactic information, it was found that sentence-level meaning did not appear crucial to perceptual learning. Specifically, listeners familiarised with syntactic prose sentences—grammatically correct sentences with real words but no sentence-level meaning (e.g., *the effect supposed to the consumer*)—achieved similar perceptual learning effects as those of listeners presented with semantically coherent English sentences. While this was the case, syntactic content alone did not appear to be the critical element behind perceptual learning. Listeners who were presented with jabberwocky sentences—sentences with real English function words but nonword content words (e.g., *the tekeen garung to the sumeeun*)—exhibited significantly less perceptual learning than listeners trained with sentences containing only real words. It was concluded that lexical information drove perceptual learning of noise-vocoded speech (Davis, et al., 2005). However, in a subsequent study, both word and nonword familiarisation material facilitated improved word recognition of noise-vocoded speech when training stimuli compromised of individual words, as opposed to the sentence-level stimuli previously employed (Hervais-Adelman, Davis, Johnsonrude, & Carlyon, 2008). Thus, lexical information may not be crucial to the facilitation of a perceptual learning response when the stimuli, as is the case with single words, can be accurately retained in short-term memory.

1.4.1.2 Familiarisation Conditions

A second issue relates to the provision, or otherwise, of feedback to augment the auditory stimuli during familiarisation. That is, whether knowledge of the atypical productions is required for perceptual learning outcomes to be realised. The evidence on this issue is varied. McQueen et al. (2006) demonstrated that learning to categorise an ambiguous phoneme could be achieved with a simple auditory listening experience (passive familiarisation). However, other studies have demonstrated that learning may necessitate more explicit conditions, wherein listeners are provided with feedback about classification performance or written information regarding the intended lexical targets (e.g., Davis, et al., 2005; Fenn, Nusbaum, & Margoliash, 2003). Learning of synthetic speech has been reported following passive experience with auditory stimuli (e.g., Koul & Hester, 2006; Reynolds, Isaacs-Duvall, & Haddox, 2002) and in studies that have employed a more explicit familiarisation procedure (e.g., Greenspan, et al., 1988; Reynolds, et al., 2002; Schwab, Nusbaum, & Pisoni, 1985). Studies comparing passive and explicit learning conditions with noise-vocoded speech have reported superior performance outcomes when the degraded stimuli is supplemented with undistorted (auditory or written) versions of the spoken targets (e.g., Davis, et al., 2005; Loebach, Pisoni, & Svirsky, 2010). In sum, it appears that perceptual learning may take place automatically when the learning entails subtle adjustments to an existing phonetic category distinction (e.g., Norris, et al., 2003). However, adaptation to an entirely novel category distinction (e.g., Logan, Lively, & Pisoni, 1991) or to an acoustic signal with substantial acoustic degradation (e.g., Davis, et al., 2005) may require more explicit familiarisation.

1.4.1.3 Amount of Familiarisation

The amount of familiarisation listeners are afforded has also varied substantially across studies. Extremely rapid learning effects have been observed following less than one minute of familiarisation with natural changes in speech rate (e.g., J. L. Miller, 1981; J. L. Miller & Liberman, 1979) and spectral degradations (e.g., Summerfield, et al., 1984; Watkins, 1981). Several minutes of familiarisation enabled perceptual learning of time-compressed (Mehler et al., 1993; Pallier, et al., 1998) and foreign-accented speech (Bradlow & Bent, 2008; Clarke & Garrett, 2004); whereas, 25 minutes (Davis, et al., 2005), nine 20 minute sessions (Rosen, Faulkner, & Wilkinson, 1999), and four sessions of one to two hours (Stacey & Summerfield, 2007) of familiarisation has been observed for learning to better recognise noise-vocoded speech. Similar to the speculations made with learning conditions, as speech becomes increasingly degraded, longer periods of familiarisation may be required for perceptual learning outcomes to be realised. While there is no conclusive evidence regarding the amount of familiarisation needed to achieve learning, studies to date would suggest that learning occurs relatively quickly, even for severely distorted speech.

1.4.1.4 Longevity of Learning

It appears that once learning has occurred, it can remain stable over a period of time. Eisner and McQueen (2005) observed learning to categorise an ambiguous phoneme remained robust following a 25 minute time lapse. This learning was evident even when additional spoken input (not containing the ambiguous phoneme) was presented during the delay. Learning effects were also reported following a lapse of 12 hours and moreover, were not dependent upon the opportunity for consolidation during sleep (Eisner & McQueen, 2005). In contrast, studies using synthetic speech have reported the need for sleep to maintain learning effects over a 12-hour period (Fenn, et al., 2003). Robust perceptual learning outcomes, measured in terms of vowel, consonant, word, and sentence recognition were observed 7-15 days for listeners familiarised with noise-vocoded speech (McGettigan, Rosen, & Scott, 2008), and improved word recognition of synthetic speech was observed at a six months follow-up test task (Schwab, et al., 1985). While limited in terms of study numbers, preliminary evidence suggests that perceptual learning may not simply be a temporary adjustment to the listener's perceptual system. Rather, learning of the unusual regularities within the acoustic signal may be long-lasting and facilitates permanent perceptual change.

1.4.1.5 Generalisation of Learning

Studies have also demonstrated that perceptual learning effects can generalise between lexical items (Davis, et al., 2005; Francis, et al., 2007). McQueen et al., (2006) and Norris et al. (2003) observed detectable changes in the categorisation of an ambiguous phoneme in words that differed from the targets encountered during the familiarisation task. This learning transfer was taken as evidence that learning may transpire at the sublexical level. Generalisation of learning to untrained words has also been reported in the recognition of foreign-accented speech (e.g., Clarke & Garrett, 2004), noise-vocoded speech (e.g., Davis, et al., 2005) and synthetic speech (e.g., Francis & Nusbaum, 2000). Such findings further support the notion that perceptual representations may be modified, at least to some degree, at the level of the phonetic unit.

While the evidence for learning transfer across novel lexical targets is relatively robust, the support for cross-speaker generalisation is less conclusive. Eisner and McQueen (2005) found that perceptual learning of an ambiguous fricative did not generalise to a novel speaker (i.e., one not included in the training condition). In contrast, Kraljic and Samuel (2006) reported cross-speaker generalisation for perceptual learning of an ambiguous stop phoneme. That phoneme learning generalised across speakers in some situations, but not in others, may indicate variations in the amount of speaker-specific information afforded by particular phoneme productions (Kraljic & Samuel, 2006). Evidence of learning transfer across speakers has also been found in studies with foreign-accented speech (Bradlow & Bent, 2008; Weill, 2001) and time-compressed speech (Dupoux & Green, 1997; Kouider & Dupoux, 2005), when the speakers exhibit similar speech patterns (i.e., speech modified in the same manner). Finally, learning of vocoded speech has been found to generalise between acoustic characteristics (Dahan & Mead, 2010; Hervais-Adelman, Davis, Taylor, Johnsrude, & Carlyon, 2011). While complete learning was achieved between different frequency regions (low-pass and high-pass filtered signals), carry-over was limited between different carrier signals (noise bands, sine waves, and pulse trains) (Hervais-Adelman, et al., 2011) and stimuli with minimal phonetic similarity (Dahan & Mead, 2010). Taken together, the findings indicate that the ability and extent to which learning can be generalised may be dependent on the acoustic similarity between the training and testing material.

1.4.2 Perceptual Learning of Dysarthric Speech

As the preceding discussion has established, there is substantial evidence regarding the perceptual benefit of prior experience with an atypical speech signal. It appears that listeners can learn to better process speech that is initially difficult to understand. While the source of learning continues to require further investigation, it is presumed that the listener learns something about the perceptual regularities of the unusual speech and can apply this information during subsequent encounters with the same signal. However, it is difficult to directly adopt the knowledge base and presumptions generated from studies using healthy speech variants (non-native) or laboratory manipulated speech (e.g., time-compressed or noise vocoded) to the perceptual learning of dysarthric speech. Unlike those forms of atypical speech, the speech degradation that occurs in individuals with neurological impairment is, by its nature, far from consistent. Speakers may deal with issues such as fluctuating muscle tone, inadequate respiratory support that worsens with fatigue, phonatory instability, and overarching deficits in articulatory movement coordination (Duffy, 2005). Thus, while some acoustic features (e.g., hypernasality) may be consistent and pervasive in a person’s speech, others may vary widely (e.g., irregular articulatory breakdowns or variable speech rate).

To date, only a handful of studies have examined listener processing and changes to speech recognition for listeners familiarised with dysarthric speech. These are reported in Table 1.1. However, given the clinical significance of improving a listener’s ability to process dysarthric speech (see section 1.2.3.3), research that investigates perceptual learning of neurologically degraded speech is critical. The majority of the existing studies have been largely clinically-based and their findings equivocal. While some research has observed significant intelligibility gains for listeners familiarised with dysarthric speech (e.g., D'Innocenzo, et al., 2006; Hustad & Cahill, 2003; Liss, et al., 2002), others have not (e.g., Garcia & Cannito, 1996; Yorkston & Beukelman, 1983). Substantial variations in research designs limit the degree to which studies can be compared; however, they do provide valuable insight into variables that may influence the nature of perceptual learning with the dysarthric signal. The following section summarises the body of research presented in Table 1.1 with regards to source of learning and the variables that appear most salient in promoting improved recognition of dysarthric speech. Implications for future research are also highlighted.

Table 1.1

Summary of the Experimental Studies that have Investigated Perceptual Learning of Dysarthric Speech

Study	Speaker Participants	Listeners Participants	Experimental Groups	Familiarisation Conditions	Familiarisation Stimuli	Transcription Stimuli	Primary Findings
D'Innocenzo, Tjaden, & Greenman (2006)	One individual with moderate mixed spastic-flaccid dysarthria secondary to traumatic brain injury.	Total of 120 normal hearing naïve individuals.	Assigned to one of 12 groups (n = 10) : various combinations of three familiarisation conditions (none, word list, paragraph) and four speaking conditions.*	Explicit.	Paragraph: Grandfather passage, or Word list: comprised of words in the Grandfather passage presented in random manner.	15 AIDS sentences.	Significantly higher intelligibility scores for listeners familiarised with either word lists or paragraph stimuli, as compared to unfamiliarised listeners. Average magnitude of difference of 10% (word list) and 8% (paragraph). No significant difference in intelligibility scores of listeners familiarised with word list stimuli compared to listeners familiarised with paragraph stimuli.
Hustad & Cahill (2003)	Five individuals with a mixed dysarthria secondary to cerebral palsy: mild hyperkinetic, mild spastic, mild spastic, severe spastic, and severe mixed spastic-hyperkinetic.	Total of 100 normal hearing naïve individuals.	Assigned to one of 5 speaker groups (n = 20): stimuli produced by one of the five speakers. NB: intelligibility scores compared across trails.	Passive.	40 HINT phrases: produced by a single speaker and presented in four sequential trials of 10 phrases.	Familiarisation phrases transcribed at time of presentation.	Significantly higher intelligibility scores across four trials for all five listener groups. Average magnitude of difference of 11%. Significant intelligibility gains for severe dysarthria were realized only between the first and third or first and fourth trials. Significant intelligibility gains for mild dysarthria were realized only between the first and second trials (no change between subsequent adjacent trials).

Liss, Spitzer, Caviness & Adler, (2002)	Twelve individuals with a moderate-severe dysarthria: six hypokinetic dysarthria and six ataxic dysarthria.	Total of 80 normal hearing naive individuals and 40 normal naïve individuals.	Assigned to one of two familiarisation groups (*n* = 40): hypokinetic or ataxic stimuli. Results compared with two control groups (*n* = 20): no familiarisation.	Explicit.	18 phrases: three per speaker.	60 phrases: 10 per speaker (one dysarthria type) + 20 phrases (other dysarthria type) i.e., 60 phrases hypokinetic speech followed by 20 phrases ataxic speech.	Significantly higher intelligibility scores for familiarised listeners compared to nonfamiliarised listeners. Average magnitude of difference of 5% (hypokinetic) and 8% (ataxic). Subset of 20 low-intelligibility phrases produced by same speech type (specific familiarisation) reflected most robust improvements. Average magnitude of difference of 16% (hypokinetic) and 21% (ataxic). Subset of 20 low-intelligibility phrases produced by other speech type (general familiarisation) reflected significant improvements compared to nonfamiliarised listeners (although gains were significantly less than specific familiarisation). No significant difference in lexical boundary error patterns for familiarised listeners compared to nonfamiliarised listeners.
Spitzer, Liss, Caviness & Adler (2000)	Twelve individuals with a moderate-severe dysarthria: six hypokinetic dysarthria and six ataxic dysarthria.	Total of 34 normal hearing naïve individuals.	Assigned to one of two familiarisation groups (*n* = 10): hypokinetic speech or ataxic speech. Results compared with two control groups (*n* = 14): no familiarisation.	Explicit.	18 phrases: three per speaker.	60 phrases: 10 per speaker (produced by same speech type encountered in familiarisation). Created by the investigators to enable error patterns to be analysed.	Significantly higher intelligibility scores for familiarised listeners compared to nonfamiliarised listeners. Average magnitude of difference of 10% (hypokinetic) and 17% (ataxic). Significantly less substitution errors for listeners familiarised with ataxic speech compared to nonfamiliarised listeners. No significant difference in substitution errors for listeners familiarised with hypokinetic speech compared to nonfamiliarised listeners.

Tjaden & Liss (1995a)	One individual with moderate-severe mixed spastic-ataxic dysarthria secondary to Cerebral palsy.	Total of 30 normal hearing naïve individuals.	Assigned to one of two familiarisation groups (*n* = 10): word list or paragraph stimuli. Results compared with a control group (*n* = 10): no familiarisation.	Explicit.	Paragraph: 12 six-word sentences presented twice, or Word list: comprised of 72 words in the paragraph presented twice in random manner.	48 phrases: 16 questions; 16 declaratives; 16 imperatives. Created by the investigators to sample a variety of phonemes and prosodic detail.	Significantly higher intelligibility scores for familiarised listeners compared to nonfamiliarised listeners. Average magnitude of difference of 5% (word list) and 9% (paragraph). No significant difference in intelligibility scores for listeners familiarised with paragraph stimuli compared to listeners familiarised with word lists.
Tjaden & Liss (1995b)	Same speaker as per Tjaden & Liss (1995a).	Total of 30 normal hearing naïve individuals.	Assigned to one of two groups (*n* = 10): treatment (speaker-oriented breath-group strategy) or treatment + familiarisation. Results compared with a control group (*n* = 10): no familiarisation with habitual speech.*	Explicit.	12 phrases: created by the investigators to sample a variety of phonemes produced in habitual speech.	48 phrases: as per Tjaden & Liss (1995a).	Significantly higher intelligibility scores for familiarised listeners compared to nonfamiliarised listeners. Average magnitude of difference of 15%. Significantly higher intelligibility scores for familiarised listeners compared to the treatment group. Average magnitude of difference of 9%.
Garcia & Cannito (1996)	One individual with severe flaccid dysarthria secondary to stroke.	Total of 96 normal hearing naïve individuals.	Assigned to one of three groups (*n* = 32): audio, visual, or audio-visual, under varying conditions: familiarisation, gesture, predictive stimuli, or situational contexts.*	Passive.	Short sample conversational speech.	16 phrases: eight “high” and eight “low” predictive.	No significant difference in intelligibility scores for familiarised listeners compared to nonfamiliarised listeners.

Yorkston & Beukelman (1983)	Nine individuals with dysarthria of varying severity levels.	Total of nine individuals (five speech pathologists and four student clinicians).	Assigned to one of two familiarisation groups (*n* = 3): passive or explicit. Results compared with a control group (*n* = 3): no familiarisation.	Passive or explicit.	Sentence list presented three times.	Novel sentence list.	No significant difference in intelligibility scores for familiarised listeners compared to nonfamiliarised listeners.

Note. Relevant studies were identified by electronic databases searches of PsycINFO, MEDLINE, CINAHL, and PubMED. The searches comprised of keywords (e.g., perceptual learning, familiarisation, adaptation) paired with the term dysarthria. In addition to these electronic searches, hand searches of studies cited within an article were conducted. From this large search, those citations in which listeners were familiarised with dysarthric speech were abstracted by the first author in Table 1.1. *In studies where additional research questions are investigated, only relevant information is reported; “passive” conditions refer to familiarisation with the dysarthric signal; “explicit” conditions refer to familiarisation with the dysarthric signal and supplementary written information of the auditory targets. Intelligibility scores = word recognition accuracy; “naïve” refers to listeners with minimal or no prior experience with dysarthria; AIDS = Assessment of Intelligibility of Dysarthric Speech (Yorkston & Beukelman, 1981); HINT = Hearing in Noise Test (Nilsson, Soli, & Sullivan, 1994).

1.4.2.1 Source of Learning

Traditionally, the dysarthrias are categorised by both type and severity, dependent upon the presence of perceptual errors (segmental goodness) and patterns (e.g., speech rate and prosody, phonatory characteristics), and the degree to which these errors and patterns impact the integrity of the acoustic signal (see section 1.2.2). This conceptualisation has motivated the majority of studies of perceptual learning in dysarthria, wherein a wide variety of dysarthria types (flaccid, spastic, ataxic, hypokinetic, hyperkinetic, spastic-flaccid, spastic-hyperkinetic and spastic-ataxic) and severities (ranging from mild to severe) have been employed. Further, the few studies that have sought to identify a source of learning (i.e., "what is learnable?") have approached dysarthric speech signal characteristics in terms of segmental versus suprasegmental degradation.

The first attempt to address "what is being learned" in a case of dysarthria was conducted by Tjaden and Liss (1995a). A non-native English speaking woman with cerebral palsy and a moderate-to-severe spastic-ataxic dysarthria provided the speech material. Normal hearing listeners transcribed her speech after first being familiarised with either her production of a read passage or with all of the words of the passage presented as a single read word list. It was expected that experience with the segmental and suprasegmental features in the read passage would be superior for perceptual learning than the single words, but ultimately both conditions benefitted intelligibility to the same degree beyond a control condition. Additional analysis confirmed that listeners learned the non-native English regularities, such as substituting /l/ for /r/.

In subsequent work, Liss and colleagues (2002) attempted to develop dependent variables that would distinguish learning about segmental regularities from suprasegmental regularities. The authors examined the LBE patterns (errors that reflect a reliance on syllable stress contrasts to inform processes of lexical segmentation) of listeners familiarised with either ataxic or hypokinetic dysarthria. While all listeners made the anticipated post-familiarisation intelligibility gains, LBE findings revealed no significant difference in error patterns made by familiarised listeners when compared with same signal transcriptions from nonfamiliarised listeners. It is possible that this result indicates that familiarisation does not improve a listener's ability to perceive differences in syllable stress contrasts with ataxic or hypokinetic dysarthria. However, it is also possible that the familiarisation procedure employed by the study, just 18 phrases, was too brief to facilitate detectable changes to the cognitive-perceptual processes of lexical segmentation.

In a post-hoc exploration of these data, Spitzer et al. (2000) completed a segmental error analysis of the listener transcripts of participants who received explicit familiarisation using phrases produced by speakers with either ataxic or hypokinetic dysarthria. The study observed changes to segmental error patterns for listeners familiarised with ataxic speech but not for those familiarised with hypokinetic speech. Listeners who heard and transcribed ataxic stimuli produced a higher proportion of target consonants in word substitutions and a lower number of substitution errors that were not phonemically related to the intended targets compared to listeners who simply transcribed the ataxic speech stimuli. Interestingly, this segmental level benefit was not evident in listeners who heard and transcribed hypokinetic speech. Absence of segmental level changes for listeners familiarised with hypokinetic speech generates the hypothesis that the source of learning may be dependent upon type of dysarthria (Spitzer, et al., 2000). However, the single-level analysis employed and, again, the fleeting familiarisation procedure, must be considered.

To date, previously published studies with dysarthric speech have yet to consider indexical properties of the signal as a potential learning source. Thus, in order to provide a more complete picture of the source of learning associated with improved recognition of neurologically degraded speech, large scale studies that examine both segmental and suprasegmental processing following a more extensive familiarisation procedure are required, as are studies that investigate the role of indexical information in perceptual learning of dysarthric speech. Such knowledge is not only key to a theoretical framework of perceptual learning of the neurologically degraded signal, but may further inform current models of perceptual processing with typical and atypical speech.

1.4.2.2 Signal Characteristics

As with source of learning, intelligibility gains for listeners familiarised with dysarthric speech may depend on the type and quality of the speech stimuli. Hustad and Cahill (2003) observed immediate improvements in recognition of mildly dysarthric speech for listeners familiarised with just 10 phrases of the speech. However, at least 30 familiarisation phrases were required for intelligibility gains to be realised with severely dysarthric speech. Consistent with these findings, Garcia and Cannito (1996) failed to report any intelligibility benefit for listeners who received a single 16 phrase familiarisation experience with severe dysarthria. Thus, it could be hypothesised that learning to better understand severely degraded dysarthric speech may necessitate greater amounts of familiarisation than that required to achieve learning of mild forms of dysarthria.

When intelligibility scores of ataxic and hypokinetic speech stimuli were matched, Liss and colleagues (2002) found that perceptual benefits of familiarisation were greatest for listeners who heard and transcribed phrases produced by the speakers with ataxic dysarthria. This suggests that the perceptual presentation of ataxic dysarthria may be more amenable to learning than that which characterises hypokinetic dysarthria. Taken together, the small number of studies conducted thus far, have demonstrated that perceptual learning may be highly dependent upon the characteristics of the signal to be learned. While further investigation into the influence of signal type and severity is warranted, signal characteristics must also be controlled for to enable valid conclusions regarding perceptual learning of dysarthric speech to be realised.

1.4.2.3 Familiarisation Conditions

To date, two types of familiarisation conditions have been employed in studies that have examined perceptual learning of dysarthric speech: passive familiarisation (degraded signal only), and explicit familiarisation (degraded signal and written transcripts of the target stimuli). Some studies that have employed passive familiarisation have reported intelligibility gains for familiarised listeners (Hustad & Cahill, 2003); whereas others have observed no perceptual benefit following a simple auditory experience with the degraded signal (Garcia & Cannito, 1996; Yorkston & Beukelman, 1983). Similarly, when studies have utilised explicit familiarisation involving supplementary written information, intelligibility gains have been documented in some studies (D'Innocenzo, et al., 2006; Liss, et al., 2002; Spitzer, et al., 2000; Tjaden & Liss, 1995a) but not in others (Yorkston & Beukelman, 1983). Only one study has directly compared intelligibility scores for listeners familiarised under either passive or explicit conditions (Yorkston & Beukelman, 1983). While the authors reported no significant intelligibility difference between the two condition groups, the validity of the comparisons may be questionable given that of the nine listener participants, five were speech pathologists and three were student clinicians (see section 1.4.2.4 for a discussion regarding listener familiarity).

Inconsistent conclusions regarding the benefit of passive and explicit familiarisation conditions are likely due, in part, to the varying amounts of familiarisation employed across studies. For example, listeners who failed to exhibit intelligibility improvements following passive familiarisation were familiarised with a short conversational sample (specific details not provided) of dysarthric speech (Garcia & Cannito, 1996). In contrast, Hustad and Cahill (2003) reported significant intelligibility gains for listeners familiarised with 40 phrases under passive learning conditions. From this comparison alone, it is speculated that when familiarisation is passive, a greater amount of training may be required for learning to transpire. Studies that have employed explicit familiarisation procedures indicate that the amount of training may have less impact on the perceptual benefit of familiarisation. However, a clear picture of how different conditions enhance, or otherwise, perceptual learning of dysarthric speech is yet to emerge.

The majority of previously published studies have attempted to control for the effects of passive or explicit familiarisation with dysarthric speech by comparing intelligibility scores of familiarised and nonfamiliarised listeners (e.g., D'Innocenzo, et al., 2006; Liss, et al., 2002; Tjaden & Liss, 1995a). As such, studies conducted thus far, are limited in their capacity to conclude that intelligibility improvements are a consequence of familiarisation with dysarthric speech. Rather, learning may transpire from the familiarisation experience. Current evidence of intelligibility benefits for listeners familiarised with dysarthric speech would be strengthened with future studies that include a control group, where listeners are familiarised with stimuli produced by neurologically intact speakers, age- and gender-matched to the speakers providing the dysarthric stimuli. In addition, perceptual learning is only of clinical value if functional gains can persist over time. Thus, studies are required to investigate whether intelligibility gains following passive and explicit familiarisation with dysarthric speech can be maintained following a significant time lapse.

1.4.2.4 Listener Familiarity

Studies that report intelligibility improvements for listeners familiarised with dysarthric speech have all employed listeners naïve to this type of speech degradation (e.g., D'Innocenzo, et al., 2006; Hustad & Cahill, 2003; Liss, et al., 2002). The single study that used speech pathologists and student clinicians as listener participants, failed to observe any intelligibility gains when listeners were familiarised with dysarthric speech under either passive or explicit conditions (relative to a group of nonfamiliarised listeners) (Yorkston & Beukelman, 1981). Thus, it may be speculated that the listeners in this study, presumed familiar with dysarthric speech, had already adapted to the degraded signal during previous unstructured interactions.

Experimental studies on listeners familiarised with dysarthric speech have yet to investigate the role of listener familiarity in perceptual learning of dysarthric speech. However, a longitudinal case study that examined intelligibility judgements of speech produced by an individual with a progressive dysarthria has reported preliminary evidence for the benefit of listener familiarity (DePaul & Kent, 2000). Intelligibility judgements made by the subject’s spouse, deemed the familiar listener, were consistently higher than a group of unfamiliar listeners (n = 24) when transcribing seven word lists collected over a period of 39 months. Thus, it appears that unstructured interaction with individuals with dysarthria may afford some degree of perceptual learning. Research is required to investigate the listener familiarity in perceptual learning of dysarthric speech and must continue to be controlled for this variable in future research designs.

1.4.2.5 Summary and Future Directions

Taken together, the small number of studies conducted thus far yield preliminary evidence that listeners can learn to better recognise neurologically degraded speech. Improved word recognition for listeners familiarised with dysarthric speech reveals a potentially promising avenue for future intervention—that is, employing a perceptual learning approach to address the intelligibility impairments that debilitate this population. Primarily, a perceptual learning rehabilitation approach would aim to increase intelligibility through improved signal processing for the trained listener. While ultimately treatment that targets universal verbal interactions is the gold standard, any approach that improves communicative effectiveness affords significant clinical application. Listener training for the management of dysarthria may be particularly applicable in the following instances: when signal production does not improve with existing interventions; when speaker-oriented approaches are not recommended (e.g., in the case of flaccid dysarthria associated with myasthenia gravis); or when co-occurring physical deficits limit the utility of augmentative or alternative approaches (e.g., communication devices, gesture, etc). Moreover, treatment that targets perceptual processes may serve as an adjunct to speaker-orientated treatment to maximise performance outcomes with particular communication partners. While such an approach may or may not afford clinical application to individuals already familiar with dysarthric speech, improving intelligibility for those communication partners unfamiliar with neurologically degraded speech, including family and friends of speakers with a recently acquired dysarthria (e.g., stroke, traumatic brain injury), holds significant value.

However, the existing literature in this area is limited. If perceptual learning is to be harnessed to build a theoretical account that supports, or otherwise, the development of listener-based treatment for the management of dysarthria, a systematic program of study grounded in current models of perceptual processing is required. The initial stages of this research should establish strong empirical evidence of intelligibility improvements, investigate the source of learning, identify optimal learning conditions and determine the longevity of learning—examination of these key components with listeners who are naïve to dysarthric speech, forms the rational and basis for the current thesis.

1.5 AIMS OF THE PRESENT THESIS

The overall aim of the current thesis was to examine perceptual learning of dysarthric speech—to build on existing empirical evidence and begin the initial steps toward the development of a theoretical framework for a perceptual learning approach to the treatment of dysarthria (see section 1.4.2.5). A moderate hypokinetic dysarthria associated with PD (described in section 1.6.1.1) was used as the test case for the programme of research. The thesis is divided into three progressive phases of research, with specific aims detailed below.

1.5.1 Phase One: Familiarisation Conditions and the Mechanisms that Underlie Improved Recognition of Dysarthric Speech.

Phase one (Chapter 3) seeks to establish the fundamentals regarding perceptual learning of dysarthric speech. This initial phase investigates the familiarisation conditions required to promote subsequent and more long-term improvements in recognition of dysarthric speech and examines the source of these intelligibility effects. The specific aims of phase one are:

1. To establish sound evidence regarding intelligibility improvements for listeners familiarised with dysarthric speech, relative to a group of control listeners familiarised with neurologically intact speech.

2. To identify whether the magnitude of intelligibility improvements for listeners familiarised with dysarthric speech is regulated by the conditions (passive versus explicit) of the familiarisation procedure.

3. To document if intelligibility improvements for listeners familiarised with dysarthric speech remain stable after a period of seven days in which no further neurologically degraded speech input is received.

4. To describe immediate and more long-term changes at segmental and suprasegmental levels of processing for listeners familiarised with dysarthric speech under passive and explicit conditions.

1.5.2 Phase Two: A Follow-up Investigation into the Mechanisms that Underlie Improved Recognition of Dysarthric Speech.

Phase two (Chapter 4) follow-ups on the phase one findings and further examines the cognitive-perceptual mechanisms that underlie improved recognition of dysarthric speech. The specific aims of phase two are:

5. To determine if the changes observed at the suprasegmental level of cognitive-perceptual processing for listeners familiarised with passage-level stimuli under both passive and explicit conditions are robust when listeners are familiarised with experimental phrases designed to heighten awareness to suprasegmental information.

6. To identify whether the magnitude of intelligibility and segmental gain for listeners familiarised with dysarthric speech are regulated by the stimuli (passages versus experimental phrase) and conditions (passive versus explicit) of the familiarisation procedure.

1.5.3 Phase Three: The Role of Linguistic and Indexical Information in Improved Recognition of Dysarthric Speech.

Phase three (Chapter 5) uses two different training tasks in order to identify the role of linguistic and indexical information in perceptual learning of dysarthric speech. The specific aims of phase three are:

7. To identify if training to attend to indexical or linguistic properties differentially affects intelligibility benefits for listeners familiarised with dysarthric speech.

8. To determine if training to attend to indexical or linguistic properties differentially affects changes at the suprasegmental and segmental level of cognitive-perceptual processing for listeners familiarised with dysarthric speech.

1.6 TEST CASE: MODERATE HYPOKINETIC DYSARTHRIA ASSOCIATED WITH PARKINSON'S DISEASE

A single type and severity of dysarthria was selected for the current series of studies. This was conducted to control for signal characteristics (see section 1.4.2.3) and also to deliver a more comprehensive investigation into perceptual learning than could have been achieved if all types and severities been examined. Moderate hypokinetic dysarthria associated with PD was selected as the test case as it provides an acoustic signal in which both segmental and suprasegmental features are compromised. The test case speakers were selected based on the presence of cardinal *speech features* of hypokinetic dysarthria to ensure relative homogeneity of the speech samples employed. The ensuing section presents a very brief overview of PD and is followed with some background on the "classic" presentation of hypokinetic dysarthria. The operational definition of hypokinetic dysarthria employed in the current research is described in Chapter two (see section 2.2.1).

1.6.1 Parkinson's Disease

Parkinson's disease is a chronic and progressive degenerative disease of the central nervous system. In his initial account of the disease, James Parkinson characterised the disorder by a resting tremor, disturbed gait, and a general slowness of movement (Parkinson, 1817). Today, tremor, rigidity, akinesia, and postural instability are considered the four cardinal features of PD (Adams, 1997). These motoric disturbances have been attributed to pathology within the basil ganglia control circuit and the degeneration of dopaminergic nigrostriatal pathways (Weiner & Lang, 1989). Further manifestations of PD are observed in its secondary motor symptoms and nonmotor symptoms, which include speech disorders, autonomic dysfunction, cognitive and neurobehavioral abnormalities, sleep disorders, and sensory abnormalities (Jankovic, 2008).

Parkinson's disease ranks as the second most common neurodegenerative disorder after Alzheimer's disease (Jankovic, 2008). Prevalence of the disorder has been estimated at 0.3% in industrialised countries, and these figures increase to 1% and 4% in individuals over 60 years and 80 years of age, respectively (Jankovic, 2008). Frequently labelled *idiopathic* PD, its cause is largely unknown. However, there appears to be a genetic predisposition to the development of PD, as well as a number of environmental factors that may increase disease susceptibility (Golbe & Langston, 1993; Weiner & Lang, 1989). Treatment of PD is largely pharmacological in nature, involving levodopa medications to reduce the motoric disturbances associated with the disease (e.g., Jankovic & Marsden, 1993; Poewe, 1993; Weiner & Lang, 1989). Additional medications, surgery, and multidisciplinary management may be employed to address the secondary symptoms. A prevalent secondary motor symptom, estimated to develop in 60 to 80% of individuals with PD (Adams, 1997), is the speech disorder hypokinetic dysarthria.

1.6.1.1 Hypokinetic Dysarthria

Hypokinetic dysarthric was first described by Grewel (1957) and Canter (1963, 1965a, 1965b) and later by Darley and colleagues (1969a, 1969b). These early reports identified monopitch, reduced stress, monoloudness, imprecise consonants, rapid speaking rate, and a harsh and breathy voice as the prominent deviant speech features of hypokinetic dysarthria. Perceptual studies since that time have reported similar findings, in addition to reduced overall intelligibility (e.g., Chenery, Murdoch, & Ingram, 1988; Darley, Aronson, & Brown, 1975; Logemann, Boshes, & Fisher, 1973; Logemann, Fisher, Boshes, & al, 1978; Ludlow & Bassich, 1983; Ludlow & Bassich, 1984; Sapir et al., 2001). The large majority of these perceptual impressions of hypokinetic dysarthria are supported by acoustic studies which have demonstrated evidence of reduced fundamental frequency variation (Canter, 1965a; Ludlow & Bassich, 1984), reduced vowel formants and trajectories (Forrest, Weismer, & Turner, 1989; Weismer, 1984), abnormal quantities of spirantization and distribution of spectral energy (Weismer, 1984), and short speech segment and transition durations (Forrest, et al., 1989; Weismer, 1984). Thus, the speech symptoms most commonly demonstrated by individuals with PD include degradation to the suprasegmental (monopitch, reduced stress, monoloudness, rapid speech rate) and segmental (imprecise articulation) properties of the signal. Audio examples of hypokinetic dysarthria of PD in American English can be found at
http://www.asu.edu/clas/shs/liss/Motor_Speech_DIsorders_Lab/Sound_Files.html.

CHAPTER TWO

Methodology

2.1 METHOD OVERVIEW

The current thesis is divided into three distinct research phases: (a) phase one: familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech (see Chapter 3); (b) phase two: a follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech (see Chapter 4); and (3) phase three: the role of linguistic and indexical information in improved recognition of dysarthric speech (see Chapter 5). While the precise nature of the perceptual learning procedures employed differs across the three phases of research, a number of methodological variables remain constant. Accordingly, the following chapter provides details on the consistent variables across the studies, with respect to listeners, speakers, acoustic analysis, experimental stimuli, and transcript analysis. Phase-specific information, including perceptual learning procedure, is detailed in the relevant chapters.

2.2 LISTENERS

One hundred and fifty healthy individuals, aged 18 to 40 years, participated in this research programme: phase one (n = 60), phase two (n = 50), and phase three (n = 40). Mean age and standard deviation of the listener groups are presented in the relevant chapters. All participants were native speakers of New Zealand English (NZE), passed a pure tone hearing screen at 20 dB HL at 1000, 2000, and 4000 Hz and at 30 dB HL for 500 Hz bilaterally, reported no experience listening to dysarthric speech (i.e., naïve listeners), and reported no identified language, learning, or cognitive disabilities. Gender was not a variable of interest; therefore, no effort was made to balance the number of male and female listeners recruited.

The majority of the listener participants were recruited from first year undergraduate classes at the University of Canterbury, Christchurch, New Zealand. Listener participants were also recruited from the author’s family and friends, local clubs and community organisations. Participants were assigned to one of the three research phases depending on time of recruitment. Individuals recruited in August to October 2009, June to July 2010 and July to August 2010, were assigned to phases one, two and three respectively. Each listener participated in one phase of the research only. Within each of the phases, participants were randomly assigned to the conditions that comprised the phase, using a computer-generated random number list.

2.3 SPEECH STIMULI

Three male native speakers of New Zealand English (NZE), with moderate hypokinetic dysarthria secondary to a primary diagnosis of Parkinson's disease (PD), and three male native speakers of NZE with neurologically intact speech (controls) provided the speech stimuli for all investigations in this thesis. Each control speaker was age-matched (within two months) to one of the three speakers with dysarthria. The speakers ranged in age from 70 to 77 years, with a mean age of 72 years. Further details of the speakers are provided in Table 2.1.

Table 2.1

Characteristics of the Speakers with Parkinson's Disease and Neurologically Intact Controls

Speakers with Dysarthria	Age	Years Post-Dx	SIT score	Control Speakers	Age
HD1	77	12	65%	CO1	77
HD2	70	11	70%	CO2	71
HD3	70	13	75%	CO3	70

Note. "HD" and "CO" refer to hypokinetic dysarthric and control speakers, respectively. The age of the HD speakers and the number of years that have elapsed since their diagnosis of Parkinson's disease (years post-dx) are presented in the first two data columns. The third data column contains the HD speaker's scores on the *Sentence Intelligibility Test* (Yorkston, Beukelman, & Hakel, 1996) as rated by one naïve listener.

2.3.1 Selection Criteria for Speakers with Dysarthria

The operational definition of hypokinetic dysarthria employed for the current study was similar to that of Liss and colleagues (1998) and derived from the Mayo Classification System (Darley, et al., 1969a; Duffy, 2005). It required that speaker participants exhibit perceptually rapid speaking rate, monopitch, monoloudness, and reduced syllable stress. All participants also exhibited a breathy and perhaps hoarse/harsh voice. In addition, speakers selected for the current studies were required to fit within a tightly constrained operational definition of a moderate intelligibility impairment—defined as a score between 65% and 75% words correct on the Sentence Intelligibility Test (SIT; Yorkston, et al., 1996). Based on this operational definition of hypokinetic dysarthria, the three participants selected for the studies exhibited highly similar segmental and suprasegmental speech characteristics.

2.3.2 Screening Procedure for Speakers with Dysarthria

An initial pool of 43 individuals with hypokinetic dysarthria was identified from neurologist recommendations and local speech-language therapy clinics as prospective speaker participants. Each of the identified individuals was posted a cover letter and an information sheet summarising the nature of the research. Ten to 12 days later, the primary researcher made telephone contact with each of the 43 potential participants. This initial telephone conversation was used to identify interest in study participation and to screen individuals for potential inclusion in the research programme. Of these, nine individuals were identified as broadly fitting the speaker selection criteria and were subsequently invited to attend a single speech assessment session at the University of Canterbury.

2.3.3 Speech Assessment Session

All speech assessment sessions were conducted in a sound-attenuated booth at the Department of Communication Disorders, University of Canterbury. Individuals were seated in a chair and fitted with a head-mounted microphone at a 5 cm mouth-to-microphone distance. Speech output elicited during the speech assessment tasks was recorded digitally to a laptop computer using *Sony Sound Forge* (v 9.0, Madison Media Software, Madison, WI) at 48 kHz (16 bit sampling rate) and stored as individual .wav files on a laptop. Samples included: (a) 15 sentences that comprised the SIT (Yorkston, et al., 1996), (b) a standard passage reading, the Rainbow Passage (Fairbanks, 1960) (see Appendix A), and (c) 72 experimental phrases (see Appendix B). Speech stimuli for the three speech tasks were presented to speakers via a PowerPoint presentation displayed on a second laptop positioned directly in front of the speakers. Large font was used to increase legibility of the stimuli. None of the nine participants experienced difficulty seeing and/or reading the speech stimuli. During the production of the stimulus phrases and passage reading, speakers were encouraged to use their 'normal, conversational' voice as they read the stimulus from a second computer monitor. Each sentence, passage reading, and phrase production, was saved to the hard drive as individual .wav files for later analysis.

2.3.4 Selection of Speakers with Dysarthria

Following the speech assessment, the PWC score on the SIT was calculated based on the transcriptions of a naïve judge (literate adult with normal hearing) who was not associated with the study. The experimental stimuli of those participants who met the operational definition of a moderate dysarthria were then analysed perceptually. Three certified speech-language pathologists experienced in the field of motor speech disorders (Stephanie Borrie, Dr Megan McAuliffe, and Dr Julie Liss) rated the experimental phrases according to perceptual impressions of the characteristics stated in the operational definition of hypokinetic dysarthria.

Of the nine speakers assessed, speech stimuli from six speakers were discarded due to the presence or absence of speech characteristics not noted during the initial telephone screening. These characteristics included the presence of a slow rate of speech, insufficient impairment of intelligibility based on SIT score, and the absence of one or more components of the operational definitions of hypokinetic dysarthria used. The remaining three speakers fit the perceptual inclusion criteria and hence were included in the study.

2.3.5 Selection of Control Speakers

Control speakers were selected according to the following criteria: (a) speakers of NZE, (b) male, and (c) age-matched to within two months to one of the three speakers with dysarthria. Potential speakers were excluded from the study if they reported any history of a neurological injury or disease, or any speech, language, hearing or voice disorder. The first three speakers approached fitted all of the selection criteria and agreed to attend a single speech assessment session at the University of Canterbury. Speech assessment sessions with the control speakers were undertaken in an identical manner to those completed with the speakers with dysarthria, with the exception that the control speakers were not required to complete the SIT[2]. All other measures and assessment conditions were kept constant.

2.3.6 Selection of Experimental Phrases

A single set of 72 experimental phrases was created by selecting 24 novel experimental phrases from each of the three speakers with dysarthria. Phrasal stress patterns were balanced, so that of the 24 phrases from each speaker, 12 were trochaic (stress on odd-numbered syllables) and 12 were iambic (stress on even-numbered syllables) in nature (see section 2.5.1, for specific details on the experimental phrases). Perceptual ratings of each phrase (see section 2.3.4) were used to ensure that each phrase included in the single speech set met the operational definition of a moderate hypokinetic dysarthria (see section 2.3.1). A second set of 72 experimental phrases was created using the corresponding control phrases produced by the neurologically intact speakers.

2.4 ACOUSTIC ANALYSIS

Acoustic analysis was performed on the two sets of experimental phrases. This process was conducted to objectively verify the presence of the perceived abnormal speech features in the speakers with hypokinetic dysarthria relative to the neurologically intact control speakers. Using *Time-Frequency Analysis* Software (TF32; Milenkovic, 2001), measures of phrase duration, fundamental frequency variation ($F0$), amplitude variation, and vowel space were calculated using standard operational definitions and procedures (Peterson & Lehiste, 1960; Weismer, 1984). These metrics were chosen to validate the presence of fast rate of speech, monotone, monoloudness and reduced syllable strength contrastivity respectively. Table 2.2 presents means and standard deviation data for phrase duration, $F0$ variation, and amplitude variation. Analysis procedures are described in further detail in the subsequent sections.

[2] The SIT was performed to ensure speakers with dysarthria all fell within a tightly constrained category of moderate intelligibility impairment. The SIT was not used to elicit experimental speech stimuli.

Table 2.2

Mean Values for each of the Acoustic Variables of Interest across the Experimental Phrases

	Mean Values (SD)	
	Dysarthric Speakers	Control Speakers
Phase Duration (ms)	1020.09 (116.46)	2031.37 (349.60)
Pitch Variation (Hz)	17.67 (4.05)	25.96 (7.23)
Amplitude Variation (dB)	6.73 (1.32)	10.92 (2.55)

2.4.1 Phrase Duration

Measures of phrase duration were obtained by placing cursors on the first and last acoustic evidence of phonemes on the spectrographic display, as per Liss et al. (1998). For initial or final voiced phonemes, cursors were placed at the first or last glottal pulse respectively. For initial or final fricatives, cursors were placed at the beginning or end of noise energy and for initial or final stop consonants, cursors were placed at the beginning or end of the burst release. Following this initial positioning of the cursors, placement remained stable for the remaining acoustic measurements performed on the phrases. An independent *t*-test, assuming equal variance, revealed that the speakers with dysarthria exhibited statistically significant reductions in phrase duration compared with the control speakers, $t(142) = 22.93$, $p < .001$, $d = 3.88$. Reduced phrase duration supported the perceptual impression of a rapid speech rate exhibited by the speakers with dysarthria.

2.4.2 Fundamental Frequency Variation

Fundamental frequency and its variation within each phrase was computed automatically using the TF32 pitch trace function key across the entire duration of the phrase. All pitch traces were inspected visually to identify and edit tracking errors, which occurred in some of the phrases produced by the speakers with dysarthria. An independent *t*-test, assuming equal variance, revealed that the speakers with dysarthria exhibited statistically significant reductions in $F0$ variation compared with the control speakers, $t(142) = 8.34$, $p < .001$, $d = 1.41$. Reduced variation in $F0$ supported the perceptual impression of monopitch exhibited by the speakers with dysarthria.

2.4.3 Amplitude Variation

Measures of amplitude and its variation within each phrase were also computed automatically over the duration of each of the phrases, using the TF32 RMS trace function. An independent *t*-test, assuming equal variance, revealed the speakers with dysarthria exhibited statistically significant reductions in amplitude variation compared with the control speakers, $t(142) = 12.36$, $p < .001$, $d = 2.06$. Reduced variation in amplitude supported the perceptual impression of monoloudness exhibited by the speakers with dysarthria.

2.4.4 Vowel Quality

To examine vowel quality, the first ($F1$) and second ($F2$) formant frequencies were measured at the temporal midpoints of six occurrences (two productions from each of the three speakers) of the vowels /i/, /a/, and /ɔ/, using both broadband spectrograms and linear predictive coding (LPC) displays. The vowels were taken from the strong syllables of the following words: *retreat*, *cheap*, *defeat*, *seat*, *meeting*, and *sheet* for /i/; *after*, *arm*, *darker*, *target*, *embark*, and *rather* for /a/; and *award*, *fortune*, *report*, *support*, *roared* and *sort* for /ɔ/. Mean formant values for each of the three vowels were used to calculate the vowel triangle area as an overall measure of vowel space for the speakers with dysarthria and matched controls. The formula was as follows: Vowel triangle area in Hz² = 0.5 x ABS [$F1_{/i/}$ x ($F2_{/a/}$-$F2_{/ɔ/}$) + $F1_{/ɔ/}$ x ($F2_{/i/}$-$F2_{/a/}$) + $F1_{/a/}$ x ($F2_{/ɔ/}$-$F2_{/i/}$)], where ABS = absolute value, $F1_{/i/}$ = first formant frequency for /i/ vowel, and so on. The vowel triangle area generated by the speakers with dysarthria was approximately 25 % smaller (171792.5 Hz^2) than the area generated by the identical vowels produced by the control speakers (233199.5 Hz^2). The perceptual impression of reduced vowel strength contrasts in the dysarthric phrases was therefore supported by the indirect measure of reduced vowel working space and the geometric area occupied by the vowel triangle derived from point vowels in strong syllables.

2.4.5 Reliability of Acoustic Measures

Twenty percent of the phrases were re-measured by the first author (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the measures of phrase duration, amplitude variation, pitch variation, and vowel formats. Discrepancies between the re-measured data and the original data are reported in terms of absolute difference. Pearson's correlation coefficients reveal the degree of association between the data sets. Table 2.3 reports strong correlations for both intra- and inter-judge reliability, with Pearson coefficients significant at, $p < .001$, for the re-analysed dependent variables.

Table 2.3.

Mean Difference and Pearson Product-Moment Correlation Coefficients for Intra- and Inter-judge Reliability of Acoustic Analysis Measures

			Intra-judge		Inter-judge	
		Speakers	*MD (SD)*	*r*	*MD (SD)*	*r*
Phrase Duration (ms)		Dysarthric	4.53 (4.31)	1.00	4.13 (5.22)	.99
		Control	5.97 (4.08)	1.00	4.89 (4.05)	1.00
Amplitude Variation (dB)		Dysarthric	0.13 (0.36)	.97	0.03 (0.08)	.99
		Control	0.03 (0.08)	1.00	0.01 (0.04)	1.00
Pitch Variation (Hz)		Dysarthric	0.27 (0.53)	.99	0.02 (0.04)	1.00
		Control	0.20 (0.42)	.99	0.06 (0.17)	1.00
Formant frequencies (Hz)	$F1$	Dysarthric	5.00 (4.55)	.99	1.25 (2.50)	1.00
		Control	6.50 (3.42)	.99	5.25 (1.89)	1.00
	$F2$	Dysarthric	15.00 (10.23)	1.00	5.50 (11.00)	1.00
		Control	8.25 (8.46)	1.00	13.00 (7.12)	1.00

Note. $F1$ = first formant; $F2$ = second format.

* $p < .001$

2.5 EXPERIMENTAL STIMULI

Primary experimental speech stimuli for the three research phases comprised a single speech set of 72 experimental phrases which had been verified perceptually and acoustically as conforming to the operational definition of moderate hypokinetic dysarthria described in section 2.3.1. In addition, readings of the Rainbow Passage (Fairbanks, 1960) from both the speakers with dysarthria and the neurologically intact controls were used as speech familiarisation stimuli in the initial phase (see Chapter 3, section 3.3.4) of the research programme.

2.5.1 Experimental Phrases

The experimental phrases were modelled on the work of Cutler and Butterfield (1992), which hypothesised that listeners rely on syllable strength to determine lexical boundaries during perception of connected speech. Each phrase consisted of six syllables and alternated phrasal stress patterns, to enable lexical boundary errors (LBEs) to be interpretable relative to syllabic strength. Half the phrases were trochaic and alternated strong-weak (SWSWSW), and the other half were iambic and alternated weak-strong (WSWSWS). The majority of the strong syllables contained full vowels and the majority of the weak syllables contained reduced vowels.

The length of the phrases ranged from three to five words. Phrases contained correct grammatical structure but no sentence level meaning (semantically anomalous) to reduce the effects of semantic and contextual knowledge on speech perception. All words were either mono- or bi-syllabic real words. The 72 phrases were used to create two speech sets of experimental phrases for the perceptual learning experiments. These were labelled speech set one and speech set two, respectively (see Appendix B).

The speech sets were balanced on a number of variables to permit direct performance comparisons, including: (a) number of phrases (36 phrases); (b) number of phrases produced by each speaker (12 phrases per speaker) (c) syllable stress pattern of the phrases (six trochaic and six iambic phrases per speaker); (d) number of words and syllables in each speech set (114 words, 216 syllables); and (e) potential number and type of lexical boundary errors (see Table 2.4). Using a Brüel & Kjær Head and Torso Simulator Type 4128-C (Brüel

& Kjær, Nærum, Denmark), all individual experimental stimuli .wav files and recordings of the Rainbow Passage (used for familiarisation material in the initial phase) were loudness calibrated to levels within ±0.1 dB. Audio presentation of all speech stimuli was set to 65 dB (A). Experimental stimuli were presented to the listeners via experimental paradigms programmed in LabVIEW 8.20 (National Instruments, TX, USA) by Dr Greg O'Beirne. The nature of these perceptual learning procedures differed for each research phase and specific details are documented in the relevant chapters.

Table 2.4

Lexical Boundary Error Opportunities by Speech Set.

LBE opportunities	Speech Set One	Speech Set Two
IS *	36	36
DS	55	54
IW	39	39
DW *	50	51

Note: "IS", "DS", "IW" and "DW" refer to lexical boundary errors defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively. * = predicted errors according to the Metrical Segmentation Strategy (Cutler & Norris, 1988).

2.6 TRANSCRIPT ANALYSIS

The total data set across the three phases of the research programme consisted of 180 transcripts of 36 experimental phrases (6480 phrases). The author independently analysed the listener transcripts for three primary measures: (1) a measure of speech intelligibility—percent words correct (PWC); (2) a measure indicative of processing at the segmental level—percent syllable resemblance (PSR); and (3) a measure indicative of processing at the suprasegmental level—presence and type of lexical boundary errors (LBEs). Percent syllable correct (PSC) were also calculated to enable the PSR measure to be viewed in context. The three primary measures are discussed in greater detail below.

2.6.1 Percent Words Correct

A PWC score, out of a total of 141 words, was tabulated for each individual listener transcript. From this, the mean PWC for the 20 listener participants in each experimental group in each research phase was derived. This score reflects a measure of intelligibility for each of the experimental conditions. Words correct were defined as those that matched the intended target exactly, as well as those that differed only by the tense “ed” or the plural “s.” In addition, substitutions between ‘‘a’’ and ‘‘the’’ were regarded as correct. The criteria for words-correct was based on other published studies which have examined listener transcripts following familiarisation with dysarthric speech (Liss, et al., 2002; Liss, et al., 1998; Liss, et al., 2000a).

2.6.2 Percent Syllable Resemblance

Transcripts were also analysed using a measure of PSR in incorrectly transcribed words. This was defined as the number of syllables that contained at least 50% phonemic accuracy to the syllable target, divided by the total number of syllable errors made. Thus, to be scored as a syllable that resembled the target, syllables with two phonemes required one correct phoneme, syllables with three phonemes required two correct phonemes, syllables with four phonemes required at least two correct phonemes, and syllables with five phonemes required at least three correct phonemes. The number of syllables that resembled the target were tallied for each transcript and divided by the total number of syllables in error for that transcript, so that the final PSR score for each transcript reflected the percentage of syllable errors that resembled the correct syllable target. Mean PSR scores for each condition were calculated. In addition, transcripts were analysed for PSC in order to examine PSR within the overall context of intelligibility. Syllables correct were defined as those that matched the intended target exactly, as well as substitutions between ‘‘a’’ and ‘‘the.’’ Each 36 phrase speech set contained a total of 216 syllables.

2.6.3 Lexical Boundary Errors

Finally, transcripts were analysed with regards to LBEs, defined as incorrect insertions or deletions of lexical boundaries. Insertion and deletion errors were further coded for location, occurring either before a strong or before a weak syllable (as per Liss, et al., 1998). Accordingly, four types of errors could be coded (see Table 2.5 for examples of coding error types): (1) insert boundary before a strong syllable (IS); (2) insert boundary before a weak syllable (IW); (3) delete boundary before a strong syllable (DS); and (4) delete boundary before a weak syllable (DW). LBE proportions for each error type were calculated as a percent score for each condition group at both initial and follow-up testing. In addition to the LBE proportion comparisons, IS/IW and DW/DS ratios based on the sum of group errors were calculated, again for each condition group at both initial and follow-up testing. According to Cutler and Butterfield (1992), these ratios are considered to reflect the strength of adherence to predicted error patterns: it is postulated that if listeners rely on syllabic strength to determine word boundaries, they will most likely make IS and DW errors. Thus, a ratio value of one reflects an equal occurrence of insertion and deletion errors before strong and weak syllables, and as the distance from one positively increases, so too does the strength of adherence to the predicted patterns of error.

Table 2.5

Example of Coding Lexical Boundary Errors from the Listener Transcripts.

Target Phrase	Listener Response	Error Type(s)
Listen final station	This is conversation	IW, DS
Afraid beneath demand	A fragment of mine	IS, DW, IS
Account for who could knock	Collect the equinox	DW, DS
For coke a great defeat	Its cooler by the sea	DW, IS
Unseen machines are green	I've seen her jeans are green	IS, IS, IS
Pick a chain for action	Flickering reaction	DW; DS, DS
Push her equal culture	Wishing he could watch her	DW; IW
Admit the gear beyond	And once again he's gone	IS, DS, IS

Note: "IS", "DS", "IW" and "DW" refer to lexical boundary errors defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively.

2.6.4 Reliability of Transcript Analysis

For each research phase, 25% of the listener transcripts were randomly selected according to a computer-generated random number list and were reanalysed by the author (intra-judge) and by a second trained judge (inter-judge) to obtain reliability estimates for the dependent variables PWC, PSR and number of LBEs. Discrepancies between the reanalysed data and the original data analysis are reported in terms of absolute mean difference and Pearson's correlation coefficients reveal the degree of association between the data sets. These values are reported in their relevant chapters.

CHAPTER THREE

Phase One: Familiarisation Conditions and the Mechanisms that Underlie Improved Recognition of Dysarthric Speech

Borrie, S. A., McAuliffe, M .J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (revision submitted). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language & Cognitive Processes: Special Edition on Speech Recognition in Adverse Conditions*.

Chapter 3 is based on the manuscript of the same name, currently under review (revision) with the Journal of Language and Cognitive Processes. Modifications to the text have been made to ensure consistency and relevance to the current chapter and thesis.

3.1 ABSTRACT

This investigation is the first in a series of three research phases which have examined perceptual learning of dysarthric speech by jointly considering intelligibility improvements and associated learning mechanisms. The current study evaluated the familiarisation conditions required to promote subsequent and more long-term improvements in perceptual processing of dysarthric speech and examined the cognitive-perceptual processes that may underlie the experience-evoked learning response. Sixty listeners were randomly allocated to one of three experimental groups and were familiarised under the following conditions: (1) neurologically intact speech (control), (2) dysarthric speech (passive familiarisation), and (3) dysarthric speech coupled with written information (explicit familiarisation). All listeners completed an identical phrase transcription task immediately following familiarisation, and listeners familiarised with dysarthric speech also completed a follow-up phrase transcription task seven days later. Listener transcripts were analysed for a measure of intelligibility, as well as error patterns at segmental and suprasegmental levels of perceptual processing. The study found that intelligibility scores for listeners familiarised with dysarthric speech were significantly greater than those of the control group, with the greatest and most robust gains afforded by the explicit familiarisation experience. Relative perceptual gains in detecting acoustic-phonetic and prosodic aspects of the signal varied dependent upon the familiarisation condition, suggesting that passive familiarisation may recruit a different learning mechanism to that of a more explicit familiarisation experience involving supplementary written information. It appeared that decisions regarding resource allocation during subsequent processing of dysarthric speech may be informed by the information afforded by the conditions of familiarisation.

3.2 INTRODUCTION

Perceptual performance can improve with experience (Volkmann, 1858) and listeners can become better able to perceive a speech signal that is initially difficult to understand (e.g., Davis, et al., 2005; Francis, et al., 2007). This experience-evoked capacity to retune or adapt the speech perception system, known as perceptual learning, is defined and described in detail in Chapter 1, section 1.4. Research with various forms of atypical speech has demonstrated that familiarisation with a less than optimal speech signal can facilitate improved recognition of the signal. While the exact source of learning remains questionable, it is commonly assumed that listeners extract regularities in the atypical acoustic pattern that facilitates or accommodates subsequent processing (see Chapter 1, section 1.4.1).

Research using healthy speech variants (non-native) or laboratory modified speech (e.g., time-compressed or noise-vocoded) provide excellent examples of this regularity, wherein segmental and/or suprasegmental aspects of these speech types vary in consistent ways. However, the acoustic degradation that characterises dysarthric speech—produced upon a platform of impaired muscle tone, inadequate respiratory support, phonatory instability, and deficient articulatory movement—frequently occurs in nonsystematic and unpredictable ways (see Chapter 1, section 1.4.2). Despite this nonsystematic variation, a small number of studies have demonstrated improved word recognition for listeners familiarised with dysarthric speech, which suggests that at least something in the dysarthric signal may be *learnable.* As summarised in Chapter 1, section 1.4.2.5, the clinical significance for perceptual learning of dysarthric speech should not be underestimated. Learning to better understand the neurologically degraded speech signal may prove key to the development of listener-focused treatments that target intelligibility impairments and hence, optimise communication success for those affected by dysarthria.

However, current experimental evidence regarding perceptual learning of dysarthric speech is limited (see Chapter 1, section 1.4.2) and further research is required if perceptual learning is to be exploited for rehabilitative gain. The present study is the first in a series that aims to investigate a listener’s capacity to improve recognition of dysarthric speech and, further, to elucidate the cognitive-perceptual source of the perceptual benefits associated with the familiarisation experience. As outlined in Chapter 1, section 1.6, moderate hypokinetic dysarthria associated with PD was targeted for the series of investigations.

While evidence of intelligibility improvements for listeners familiarised with dysarthric speech has been reported (e.g., D'Innocenzo, et al., 2006; Liss, et al., 2002), the absence of adequate experimental control has reduced the strength of reported findings. Research conducted thus far has attempted to assess the magnitude of perceptual learning effects by comparing intelligibility scores from listeners familiarised with dysarthric speech to those who have not received familiarisation. In such cases, particularly where the familiarisation material affords similarities to the test material, it is challenging to separate the perceptual improvements that result from familiarisation with dysarthric speech, to those that may arise simply from the familiarisation experience (e.g., Hustad & Cahill, 2003; Liss, et al., 2002). In order to reliably attribute perceptual benefits to familiarisation with dysarthric speech, research is required to compare learning effects from listeners familiarised with dysarthric speech to listeners familiarised with same stimuli produced by neurologically intact speakers.

A significant methodological variation across the existing research is found in the type of familiarisation conditions employed (see Chapter 1, section 1.4.2.3). There is evidence that learning may transpire automatically, as a result of passive familiarisation to the degraded auditory productions (e.g., Hustad & Cahill, 2003). There is also evidence to suggest that more explicit familiarisation involving supplementary written information may be required for perceptual benefits of familiarisation to be realised (e.g., Liss, et al., 2002). Thus far, only one study has directly compared intelligibility scores following passive versus explicit familiarisation (Yorkston & Beukelman, 1983), reporting that word recognition accuracy did not differ across these two learning conditions. The study also observed no difference in word recognition accuracy when familiarisation groups were compared to a group of nonfamiliarised listeners. However, given the nature of the nine listener participants (five speech pathologists and three student clinicians), the validity of these findings is questionable (see Chapter 1, section 1.4.2.4). As such, existing research has yet to provide conclusive evidence of the learning conditions required to establish and/or promote improved recognition of dysarthric speech.

Clinically, the perceptual benefit of a familiarisation experience is of functional value only if improvements can persist over time. Therefore, research is also required to identify whether intelligibility scores achieved immediately following experience with dysarthric speech can remain stable over a period in which no further neurologically degraded speech input is received. While studies with artificially manipulated signals have demonstrated that the perceptual benefit of a familiarisation experience can continue following a significant time lapse (McGettigan, et al., 2008; Schwab, et al., 1985), the few studies that have examined familiarisation with dysarthric speech have yet to investigate this phenomena. Bearing in mind that dysarthric speech is characterised by multiple segmental and suprasegmental distortions, improved recognition of this type of speech presumably involves a number of different processing levels and significant cognitive resource. Accordingly, investigation into the longevity of perceptual learning effects with the neurologically degraded signal holds both clinical and theoretical significance.

If significant intelligibility improvements can be realised following a familiarisation experience, a critical question remains—what is the source of this performance gain? Is it that listeners have learnt something about the global prosodic features of the speech signal and can exploit these cues more readily or rather, have they learnt to recognise regularities in phonological form and apply these to their own internal representations? The cognitive-perceptual mechanisms that underlie intelligibility improvements are currently not well understood (see Chapter 1, section 1.4.1 for full details).

Taking a traditional view of speech perception, we can hypothesise that the learnable and useful regularities in the dysarthric signal will facilitate the perceptual process of lexical segmentation, lexical activation, and/or lexical competition (see Chapter 1, section 1.3.1 for an explanation of these processes). One could imagine, for example, that prior exposure to the rapid articulation rate common in hypokinetic dysarthria may allow listeners to modify their expectations of phoneme duration which in turn, may reduce ambiguity and facilitate lexical activation and competition. Or perhaps exposure to the rapid speaking rate and reduced variation in fundamental frequency facilitates lexical segmentation by encouraging attention to alternative syllabic strength contrast cues. To date however, only two studies have begun to shed light upon the possible cognitive-perceptual changes associated with improved intelligibility of dysarthric speech (see Chapter 1, section 1.4.2.3 for full details). These studies have proposed that the performance benefits associated with a familiarisation experience may occur with improved processing of segmental information (Liss, et al., 2002; Spitzer, et al., 2000). However, evidence is limited and current findings have not led to clear answers. Further examination of both segmental and suprasegmental processing for listeners familiarised with dysarthric speech is required to provide a more complete picture of the learning mechanisms that may underlie improved intelligibility of neurologically degraded speech.

3.2.1 Current Study

The present investigation, therefore, aims to extend the existing body of literature pertaining to perceptual learning of dysarthric speech—to establish strong empirical evidence of intelligibility improvements, verify the familiarisation conditions that promote learning, document stability of learning effects over time, and investigate the source of learning. The following four questions were addressed: (1) Do listeners who are familiarised with dysarthric speech achieve higher intelligibility scores relative to listeners who are familiarised with neurologically intact speech; (2) Is there an effect of familiarisation condition, in which the magnitude of perceptual gain is regulated by the type of familiarisation experience (passive versus explicit); (3) Do perceptual gains remain stable after a period of seven days in which no further dysarthric speech input is received; (4) Are perceptual gains accompanied by changes at the segmental and/or the suprasegmental level of cognitive-perceptual processing?

3.3 METHOD

3.3.1 Research Design

A between-groups design was used to investigate perceptual learning effects associated with three different familiarisation conditions. Three groups of listeners were familiarised with passage readings under one of three experimental conditions: (1) neurologically intact speech (control), (2) dysarthric speech (passive familiarisation), and (3) dysarthric speech coupled with written information (explicit familiarisation). Following familiarisation, all listeners completed an identical phrase transcription test (initial test). Listeners familiarised with dysarthric speech returned seven days following the initial familiarisation experience and completed a second phrase transcription test (follow-up test) involving novel phrases.

3.3.2 Listeners

Data were collected from 60 young healthy individuals (47 women, 13 men) with a mean age of 25.5 years (*SD* = 5.2). See Chapter 2, section 2.2 for further details of the listener participants.

3.3.3 Speech Stimuli

Speech familiarisation material consisted of readings of the Rainbow Passage (Fairbanks, 1960) (see Appendix A) spoken by individuals with dysarthria and neurologically intact control speakers. Test material consisted of the 72 experimental phrases that made up speech set one and speech set two (see Appendix B), re-labelled for use in the current study as the *initial test speech set* and *follow-up test speech set*, respectively. See Chapter 2, section 2.5 for further details of the speech stimuli.

3.3.4 Perceptual Learning Procedure

The 60 listener participants were randomly assigned to one of three experimental groups, so that each group consisted of 20 participants. The experimental groups were labelled as follows: (a) control, (b) passive-passages, and (c) explicit-passages. The experiment was conducted in two primary phases: (1) familiarisation phase and (2) initial test phase, and the passive and explicit groups participated in a third (3) follow-up test phase. Figure 3.1 contains a diagrammatic representation of the perceptual learning procedure employed.

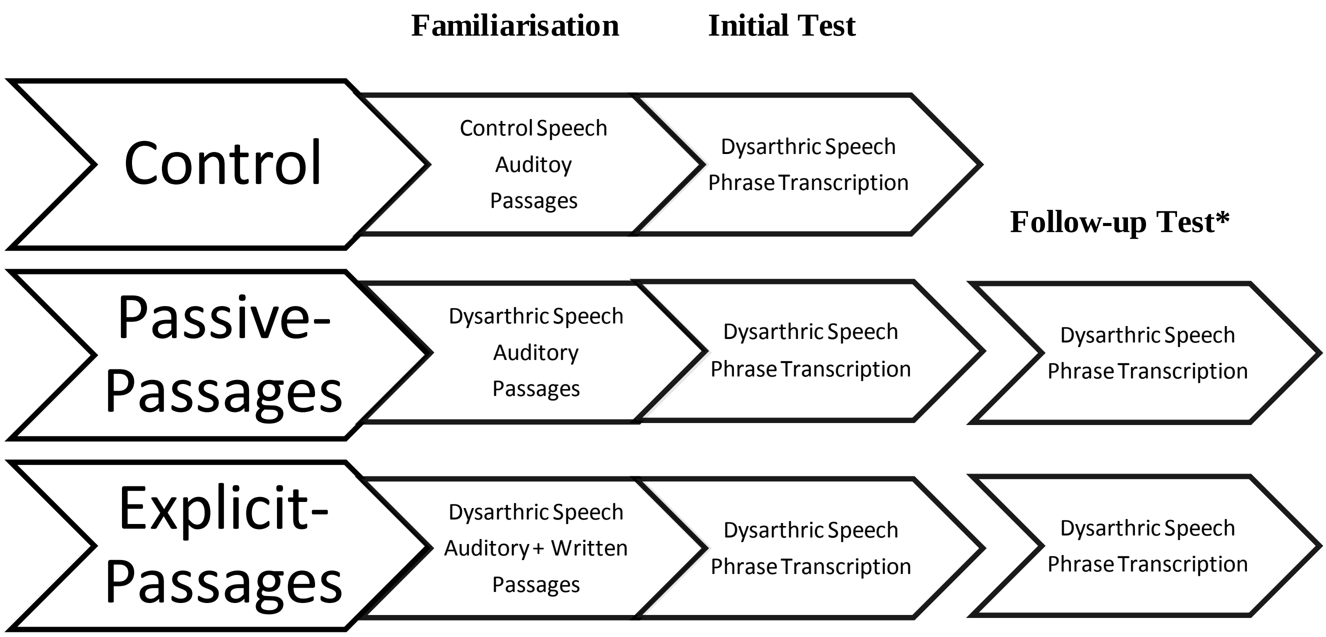

Figure3.1. Perceptual learning procedure divided into three phases. The first column provides details of the familiarisation phase, the second provides details of the initial test phase, and the third provides details of the follow-up test phase. * = conducted seven days post-familiarisation and initial test phase.

The experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 pro). Listeners were tested either individually or in pairs, located to eliminate visual distractions. The experiment was presented via a laptop computer, pre-loaded with the experimental procedure. Participants were told that they would undertake a listening task followed by a transcription task, and that task-specific instructions would be delivered via the computer programme. This process was employed to ensure identical stimulus presentation methods across participants.

During the *familiarisation* phase, listeners in the control group were presented with three readings of the rainbow passage, each produced by a different speaker with neurologically intact speech. To ensure each speaker was heard in each position a similar number of times, the order in which each of the 20 participants heard the three speakers was counterbalanced. For example, two of the speakers were heard in the first position seven times and one speaker six times, with similar ratios for the second and third positions. The order was then randomized using the Knuth implementation of the Fisher-Yates shuffling algorithm (Knuth, 1998). Participants were instructed to simply listen to the three speech samples. Listeners in the passive-passages group were also given the same instruction but were presented with three readings of the rainbow passage; each produced by a different speaker with dysarthria. Listeners in the explicit-passages group were presented with the same dysarthric stimuli as the passive-passages group, however they were provided with a written transcript of the intended targets on the computer screen and were instructed to carefully read along as they listened. The order of familiarisation material was controlled using identical procedures to that described for the control group.

Immediately following the familiarisation task, all three experimental groups participated in an identical *initial test* phase in which they transcribed the initial test speech set. Phrases were presented one at a time and listeners were asked to listen carefully to each phrase and to type exactly what they heard. Listeners were told that all phrases contained real English words but that the phrases themselves would not make sense. They were told that some of the phrases would be difficult to understand, and that they should guess any words they did not recognise. Listeners were told to place an “X” to represent part of a phrase, if they were unable to make a guess. They were given 12 seconds to type each response. Listeners in the passive-passages and explicit-passages groups were asked to return in seven days to participate in the *follow-up test* phase, in which they transcribed the follow-up test speech set. Transcription instructions at the follow-up test were identical to those received at the initial testing phase. The 36 phrases in both the initial and follow-up test speech sets were presented randomly to each of the 60 listener participants.

3.3.5 Transcript Analysis

The total data set consisted of 100 transcripts of 36 experimental phrases: 60 transcripts of the initial test speech set and 40 of the follow-up test speech set. The author independently analysed the listener transcripts for PWC, PSR, PSC, and the presence and type of LBEs. Details regarding the analysis and calculation of these measures are found in Chapter 2, section 2.6. The reliability of the transcript analysis for the 100 transcripts was then measured (details of the method of reliability measurement are reported in Chapter 2, section 2.6.4). The first author and a second trained judge reanalysed 25% of the transcripts. Discrepancies between the reanalysed data and the original data analysis are reported in terms of absolute mean difference. Pearson product-moment correlation coefficients were computed to assess the relationship between the data sets. Table 3.1 summarises the results. A strong, positive correlation between the reanalysed data and original data was found for the analysis of the transcripts.

Table 3.1

Mean Difference and Pearson Product-Moment Correlation Coefficients for Intra- and Inter-judge Reliability of the Transcript Analysis

	Intra-judge		Inter-judge	
	MD (SD)	*r*	*M (SD)*	*r*
PWC	0.31 (0.41)	.99*	0.58 (0.65)	.99*
PSR	1.20 (1.08)	.96*	1.44 (0.70)	.97*
LBE	0.88 (0.73)	.99*	1.04 (0.73)	.99*

Note. PWC = percent words correct; PSR = percent syllables correct; LBE = lexical boundary errors.

* $p < .001$.

3.4 RESULTS

3.4.1 Percent Words Correct

Figure 3.2 reflects the mean PWC scores for all three experimental groups at initial and follow-up tests. A one-way analysis of variance (ANOVA) showed a significant group effect for PWC scores immediately following familiarisation, $F(2, 57) = 89.15$, $p < .001$, $\eta^2 = .76$. *Post-hoc* tests, using Bonferroni correction, revealed that PWC scores achieved by the explicit-passages group were significantly higher than those evident in the passive-passages group, $t(38) = 5.30$, $p < .001$, $d = 1.84$, and the control group, $t(38) = 13.24$, $p < .001$, $d = 3.76$, and that PWC scores achieved by the passive-passages group were significantly higher than those evident in the control group, $t(38) = 8.09$, $p < .001$, $d = 2.66$. Thus, immediate intelligibility improvements were realised for both groups familiarised with dysarthric passages, with the greatest gains observed for the listeners familiarised under explicit conditions.

Paired *t*-tests were used to examine the within-group stability of intelligibility gains over time by comparing PWC scores from the initial and follow-up tests. Comparisons revealed that the PWC scores for both the passive-passages group, $t(19) = 13.94$, $p < .001$, $d = 3.72$, and the explicit-passages group, $t(19) =12.48$, $p < .001$, $d = 2.47$, declined significantly over the seven day interval. When PWC scores from the passive-passages and explicit-passages groups at follow-up were compared with the control group, a one-way ANOVA revealed a significant group effect, $F(2, 57) = 11.99$, $p < 0.001$, $\eta^2= .30$. *Post-hoc* tests, using Bonferroni correction, indicated that while the PWC scores for the passive-passages group at follow-up were similar to those evident in the control group, $t(38) = 0.53$, $p = 1.0$, $d = .19$, the PWC scores for the explicit-passages group at follow-up were significantly higher than both the control group, $t(38) = 4.48$, $p < .001$, $d = 1.22$, and the passive-passages group, $t(38) = 3.94$, $p < .001$, $d = 1.37$. Thus, while intelligibility declined over seven days for both groups familiarised with dysarthric passages, some intelligibility carry-over was observed for the listeners familiarised under explicit conditions.

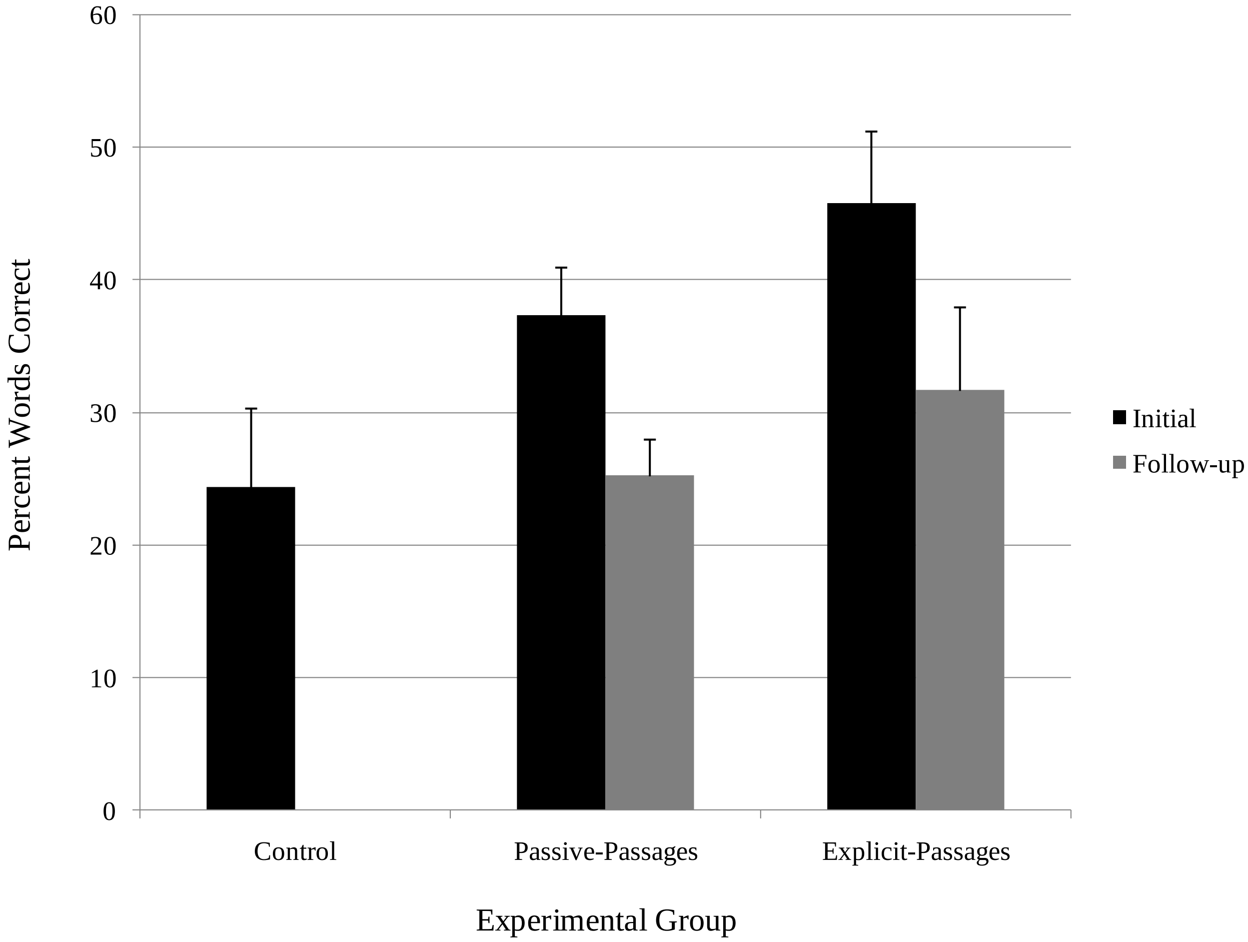

Figure 3.2. Mean percent words correct (PWC) for listeners by experimental group at the initial and follow-up tests. Bars delineate + 1 standard deviation of the mean.

3.4.2 Percent Syllable Resemblance

Figure 3.3 reflects the mean PSR scores, in addition to the mean PSC scores, for the three experimental groups at initial and follow-up tests. Pearson product-moment correlation coefficients demonstrated a strong relationship between the variables of PSC and PWC for all conditions (see Table 3.2). Accordingly, statistical analysis was performed on the PSR data only, as PSC findings are reflected in the analysis of PWC (see section 4.3.1).

A one-way ANOVA on the PSR scores revealed a significant group effect immediately following familiarisation, $F(2, 57) = 11.17$, $p < .001$, $\eta^2 = .28$. *Post hoc* tests, using Bonferroni corrections, demonstrated that PSR scores achieved by both the passive-passages group, $t(38) = 2.98$, $p = .01$, $d = 1.05$, and the explicit-passages group, $t(38) = 4.67$, $p < .001$, $d = 1.44$, were significantly higher than the control group. There was no significant difference in PRS scores achieved by the passive-passages and explicit-passages groups, $t(38) = 1.69$, $p = .29$, $d = .50$. Thus, passive familiarisation with dysarthric passages facilitated similar benefits to a segmental measure of perceptual processing as explicit familiarisation with dysarthric passages.

Paired *t*-tests were used to examine the within-group stability of segmental gains over time by comparing PSR scores from the initial and follow-up tests. Comparisons revealed that while a small increase in the PSR scores was observed at follow-up for both groups, these differences were not significant for the passive-passages group, $t(19) = 1.3$, $p = .20$, $d = .40$, and the explicit-passages group, $t(19) = 1.6$, $p = .11$, $d = .40$. When PSR scores from the passive-passages and explicit-passages groups at follow-up were compared with the control group, a one-way ANOVA revealed a significant group effect, $F(2, 57) = 20.69$, $p < .001$, $\eta^2 = .42$. *Post-hoc* tests, using Bonferroni correction, demonstrated that PSR scores achieved by both the passive-passages group, $t(38) = 4.49$, $p < .001$, $d = 1.37$, and the explicit-passages group, $t(38) = 6.24$, $p < .001$, $d = 2.18$, were significantly higher than the control group. There was no significant difference in PRS scores achieved by the passive-passages and explicit-passages groups at follow-up, $t(38) = 1.75$, $p = .26$, $d = .50$. Taken together, the within- and between-group comparisons on the PSR data show that the benefits to a measure of segmental processing for both groups familiarised with dysarthric passages remained robust over seven days.

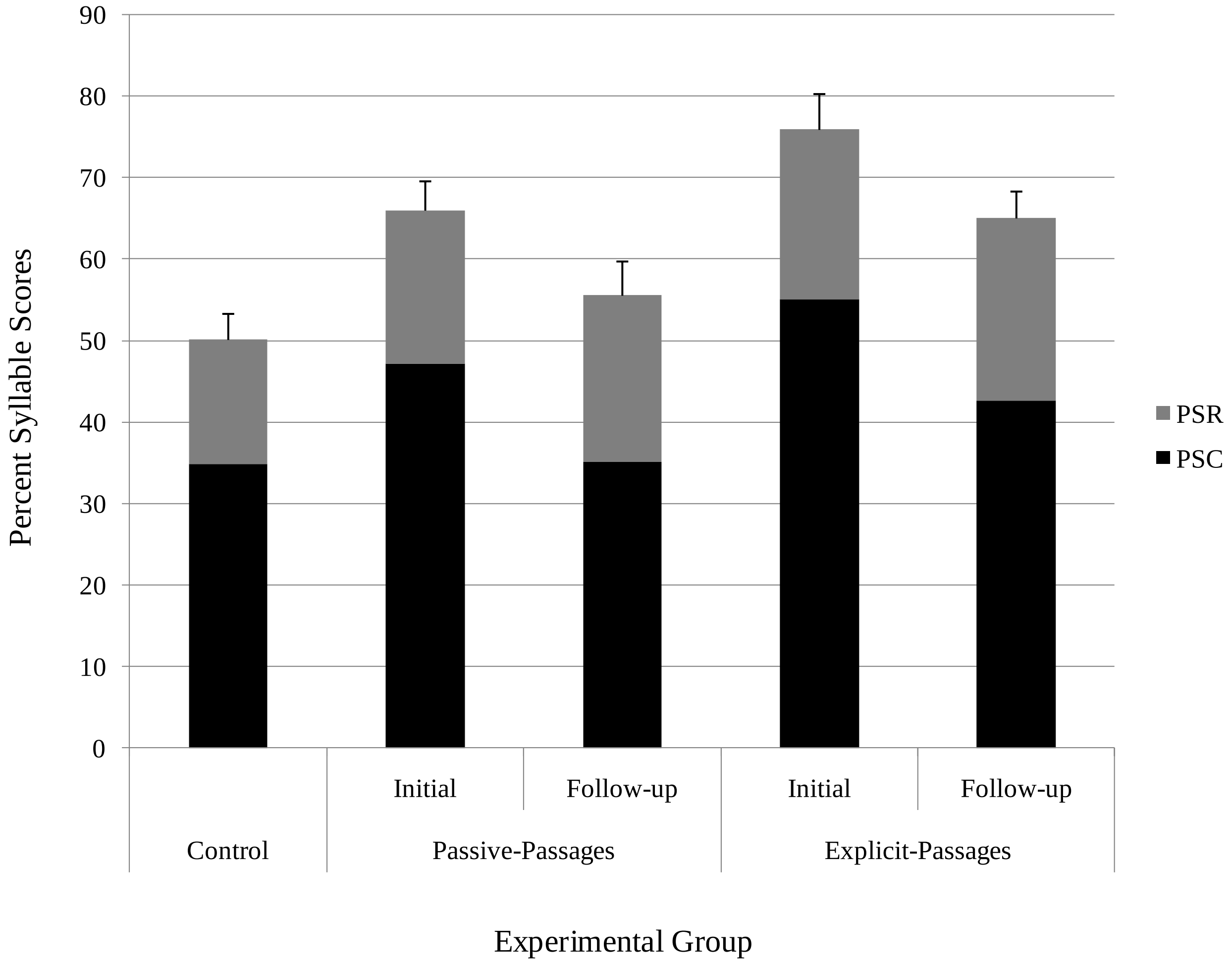

Figure 3.3. Mean percent syllable correct (PSC) and mean percent syllable resemblance (PSR) for listeners by experimental group at the initial and follow-up tests. Bars delineate + 1 standard deviation of the mean PSR data.

Table 3.2

Mean Difference and Pearson Product-Moment Correlation Coefficients between Percent Words Correct and Percent Syllables Correct for Listeners by Experimental Group

Group [a]	*MD (SD)*	*r*
Control	10.5 (2.9)	.90*
Passive-Passages	9.9 (2.1)	.88*
Passive-Passages: Follow-up	9.9 (2.9)	.71*
Explicit-Passages	9.2 (3.1)	.84*
Explicit-Passages: Follow-up	10.9 (3.7)	.81*

[a] $n = 20$

* $p < .001$

3.4.3 Lexical Boundary Error Patterns

Table 3.3 contains the LBE category proportions and the sum IS/IW and DW/DS ratios for the three experimental groups at the initial and follow-up tests. Contingency tables were constructed for the total number of LBEs by error type (i.e., insertion/deletion) and error location (i.e., before strong/weak syllable) for the groups to determine whether the variables were significantly related. A within-group chi-square analysis revealed a significant interaction between the variables of type (insert/delete) and location (strong/weak) for the data generated by the control group, $X^2(1, N = 20) = 33.15$, $p < .001$, and the explicit-passages group—both immediately following familiarisation, $X^2(1, N = 20) = 76.95$, $p < .001$, and at follow-up, $X^2(1, N = 20) = 128.27$, $p < .001$. In both the control and the explicit-passage groups, erroneous lexical boundary insertions occurred more often before strong than before weak syllables, and erroneous lexical boundary deletions occurred more often before weak than before strong syllables. Such LBE error patterns are predicted (Cutler & Butterfield, 1992 see also Chapter 2, section 2.6.3). Ratio figures reflect the strength of adherence to these predicted error patterns—the greater the positive distance from “1,” the stronger the adherence. Relative to the control group, the magnitude of the IS/IW ratio is substantially greater for explicit-passages group. This indicates that listeners familiarised with dysarthric passages under explicit conditions learnt to utilise syllabic stress contrast cues to inform speech segmentation. This finding was not evidenced in the data of the passive-passages group, at either the initial or follow-up testing. While there was a small increase in the number of erroneous lexical boundary insertions that occurred before a strong syllable relative to a weak syllable, there was a small decrease in the number of erroneous lexical boundary deletions that occurred before a weak syllable relative to a strong syllable. Differences, however, were not significant both immediately following familiarisation, $X^2(1, N = 20) = 0.22$, $p = .71$, and at follow-up, $X^2(1, N = 20) = 2.25$, $p = .14$. No significant relationship between the type and location of LBEs for the passive-passages group indicates that the listeners familiarised with dysarthric passages under passive conditions did not learn to utilise syllabic stress contrast cues to inform speech segmentation.

A between-group chi-square analysis was used to examine differences in error distribution between the three experimental groups. Results identified significant differences in error distribution between the control and passive-passages groups, $X^2(3, N = 40) = 38.98$, $p < .001$, and the passive-passages and explicit-passages groups, $X^2(3, N = 40) = 109.19$, $p < .001$. No significant difference was found between the control and explicit-passages groups, $X^2(3, N = 40) = 6.34$, $p = .10$. Thus, the relative distribution of errors observed for the control group were similar to those observed for the listeners familiarised with dysarthric passages under explicit conditions, but this error pattern was significantly different to that observed for the listeners familiarised with dysarthric passages under passive conditions.

Table 3.3

Category Proportions of Lexical Boundary Errors Expressed in Percentages and Sum Error Ratio Values for Listeners by Experimental Group

Group [a]	%IS	%IW	%DS	%DW	IS-IW Ratio	DW-DS Ratio
Control	37.15	15.84	19.55	28.21	2.4	1.4
Passive-Passages	27.31	22.69	28.41	21.59	1.2	0.8
Passive-Passages: FU	29.48	28.87	23.92	17.73	1.0	0.7
Explicit-Passages	42.42	12.31	16.70	28.57	3.5	1.7
Explicit-Passages: FU	42.12	14.95	12.06	30.87	2.8	2.6

Note: “IS”, “DS”, “IW” and “DW” refer to lexical boundary errors defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively. FU = Follow-up.

[a] n = 20

3.5 DISCUSSION

The present study provides evidence of perceptual learning for listeners familiarised with dysarthric speech and enables a number of conclusions to be drawn. First, intelligibility improved substantially following a relatively brief familiarisation experience with dysarthric stimuli. Second, the magnitude and robustness of the intelligibility benefits were influenced by the familiarisation conditions. Finally, performance gains were associated with changes in the processing of both segmental and suprasegmental aspects of the degraded signal. However the perceptual changes at these processing levels appeared to vary as a function of familiarisation condition. Such findings support a dynamic and adaptable speech perception system, which is further discussed with regards to speech intelligibility and cognitive-perceptual processing.

Significantly higher intelligibility scores were observed for listeners familiarised with dysarthric speech compared with those familiarised with control speech. Improved processing of the dysarthric signal demonstrates that listeners can learn to better understand neurologically degraded speech. This provides evidence for a dynamic model of perceptual processing that enables online adjustments to acoustic features of dysarthric speech. Key, however, is that explicit familiarisation offered superior performance gains than those afforded by passive familiarisation, as has been previously reported with perceptual learning of noise-vocoded speech (Davis, et al., 2005; Loebach, et al., 2010). In addition to significantly larger intelligibility benefits, explicit familiarisation also facilitated some intelligibility carry-over (at seven days post-familiarisation). Listeners who received passive familiarisation did not exhibit any performance gains at follow-up. From the intelligibility data alone, it would appear that passive familiarisation with the degraded signal alone is not sufficient to facilitate any long-term changes in perceptual processing. This likely reflects the fact that there was less learning in the passive condition because, based on the performance of the control group, only approximately 25% of the words in the phrases were recognisable. Even if limited, it has been speculated that the ability to recognise some words enables listeners to use acoustic-phonetic information to modify phonemic representations (e.g., Eisner & McQueen, 2005; Norris, et al., 2003). Thus, it can be speculated that the addition of the passive-passage familiarisation allowed listeners to better exploit the 25% understandable words for an additional 13% gain. Less robust learning would lead to faster decay if, as in modular theories, learning is viewed as a temporary perceptual adjustment, allowing representations to return to pre-perceptual learning parameters over time (Kraljic & Samuel, 2005).

If intelligibility scores were considered in isolation, the explanation that the performance benefit associated with passive familiarisation was simply enhanced when familiarisation was more explicit could be assumed. However, error patterns at segmental and suprasegmental levels of perceptual processing reveal that intelligibility differences between experimental groups were not simply a case of the magnitude of learning. Listeners familiarised with dysarthric speech achieved a significantly higher percentage of syllables that bore phonemic resemblance to the targets (not including correctly transcribed syllables) relative to the control group. Thus, it appears that experience with dysarthric speech enabled listeners to better map acoustic-phonetic aspects of the disordered signal onto existing mental representations of speech sounds. This finding extends support for previous studies which have postulated that improved recognition of dysarthric speech is sourced from segmental level benefits (Liss, et al., 2002; Spitzer, et al., 2000). However, it is difficult to account for the superior intelligibility benefits observed in listeners who received explicit familiarisation, given that the PSR scores were similar for both passive and explicit familiarisation conditions. Furthermore, there appeared to be relative maintenance of the segmental benefits afforded by both passive and explicit familiarisation conditions at follow-up. The PSR scores did not diminish on day seven for either of the familiarised groups. Thus, despite poorer words-correct intelligibility performance in the passive-passages group, the perceptual benefits to segmental processing appeared to remain. Given that word recognition scores returned to levels similar to that of the controls for passively familiarised listeners, robust improvements in phoneme perception at follow-up for this group are unexpected. Stable PSR scores in the face of a substantial intelligibility decline would serve to demonstrate that passive familiarisation to dysarthric speech does improve subsequent acoustic-phonetic mapping at seven days following the exposure experience. If the measure of syllabic resemblance is a valid index of phoneme perception accuracy, these findings raise the possibility that learning decay may occur at different rates across different levels of analysis. However, it is also possible that the decay in word recognition scores may, to some degree, be influenced by the amount of familiarisation listeners receive. While the quantity of familiarisation material was substantially more than the amount that is generally employed to study this phenomenon (e.g., D'Innocenzo, et al., 2006; Liss, et al., 2002; Tjaden & Liss, 1995a), whether increased periods of familiarisation would facilitate more robust intelligibility benefits provides a valuable direction for future investigations.

Another unexpected finding calls into question the conclusion that the difference between passive and explicit familiarisation simply reflects how much the listener has learnt. Comparison of the LBE error patterns of the control and explicit-passages groups reveal expected results. Both groups made significantly more predicted (IS and DW) errors than non-predicted (IW and DS) errors, a pattern which conforms to the MSS hypothesis (see Chapter 2, section 2.6.3 for detailed explanation). Furthermore, this pattern was stronger for the group that received explicit familiarisation than for the control group. While reduced syllable stress contrasts are a cardinal feature of hypokinetic dysarthria (Darley, et al., 1969a), the presence of written information during experience with the degraded signal presumably enabled listeners to learn something about the reduced and aberrant syllabic stress contrast cues by drawing attention to relevant acoustic information (e.g., Goldstone, 1998; Nosofsky, 1986). Such findings are supported by evidence that listeners relied on syllabic stress information to facilitate lexical segmentation of speech produced by individuals with hypokinetic dysarthria (Liss, et al., 1998), although a relatively small familiarisation procedure in a subsequent study did not elicit significant changes in LBE error patterns (Liss, et al., 2002).

The unexpected finding, then, comes with the analysis of the passive familiarisation LBE data. This group appeared to largely ignore syllabic strength contrast cues to inform speech segmentation. In contrast to listeners in the control and explicit-passages groups, listeners who received passive familiarisation were just as likely to make unpredicted errors (IW and DS) as they were to make predicted errors (IS and DW) (see Chapter 2, section 2.6.3 for explanation of MSS error patterns). This is a remarkable finding given that the sole difference between the passive and explicit groups was the addition of written information for listeners familiarised with dysarthric speech under explicit conditions. Furthermore, similar LBE patterns were observed for both passive and explicit groups at follow-up suggesting, perhaps, the persistence of cognitive-perceptual strategies that were engendered by each familiarisation procedure. Thus, LBE data reveals that familiarisation conditions may differentially influence learning of suprasegmental properties. The presence of written information regarding the lexical targets appeared to promote syllabic stress contrasts as an informative acoustic cue, whereas experience with degraded signal alone essentially eliminated any cognitive attention toward this prosodic information. Interestingly, this conclusion appears to be at odds with some of the perceptual learning literature that has speculated on conditions required to achieve learning. Research has identified that perceptual learning of a signal in which segmental properties have been artificially manipulated (e.g., noise-vocoded speech) may depend on knowledge of the lexical targets (e.g., Davis, et al., 2005), whereas improved recognition of a signal in which the suprasegmental information has been modified (e.g., time-compressed speech) has been reported in the absence of any supplementary information regarding the degraded productions (e.g., Pallier, et al., 1998; Sebastian-Galles, et al., 2000). Future studies are needed to investigate why, with the neurologically degraded signal, segmental properties appear to be learned relatively automatically and yet attention toward suprasegmental information may necessitate more explicit learning conditions. In addition, research with other types and severities of dysarthric speech will enable a more comprehensive picture of perceptual learning processes to be established.

3.6 CONCLUSION

The current study yields empirical support for perceptual learning of dysarthric speech. There is evidence to suggest that greater and more robust performance gains are achieved when the degraded signal is supplemented with written information under explicit learning conditions. However, there is also evidence to suggest that, for this particular pattern and level of speech degradation, the learning afforded by passive familiarisation may be qualitatively different to that afforded by explicit familiarisation. Thus, the current study has revealed a possible relationship between familiarisation conditions (passive verses explicit) and subsequent processing of dysarthric speech. Further research is, however, required to validate such a speculation.

CHAPTER FOUR

Phase Two: A Follow-up Investigation into the Mechanisms that Underlie Improved Recognition of Dysarthric Speech

4.1 ABSTRACT

Chapter 3 reported that the intelligibility benefits afforded by a familiarisation experience with dysarthric speech were superior when the degraded speech signal was supplemented with written information. Discrepancies were also evident in speech segmentation strategies, revealing that performance differences were more than simply magnitude of benefit. It was speculated that the learning afforded by passive familiarisation may be qualitatively different to that which occurs with explicit familiarisation. To follow up on this finding, the current study aimed to determine if the key variable behind the use of particular segmentation strategies was simply the presence or absence of written information during familiarisation. Forty listeners were randomly assigned to a passive or explicit condition group (as per Chapter 3) and were familiarised with experimental phrases designed to heighten awareness of alternating syllabic stress patterns. Immediately following familiarisation, all listeners completed an identical phrase transcription task. The resultant data were compared to corresponding data from Chapter 3, wherein listeners were familiarised with a short passage reading under either passive or explicit conditions. The present study found that listeners familiarised with phrases under passive or explicit conditions demonstrated similar segmentation strategies of exploiting syllabic stress contrast cues to inform lexical boundary decisions. Thus, it was concluded that segmentation strategies are not merely influenced by the presence or absence of written information during familiarisation. In addition, intelligibility data revealed that performance improvements were greatest when linguistic properties of the degraded speech signal were emphasised with written information or linguistically-predictive familiarisation stimuli. Taken together, the findings suggest that perceptual learning of dysarthric speech is influenced differentially by the information afforded within the familiarisation procedure.

4.2 INTRODUCTION

Fundamental to an understanding of spoken language recognition is the ability to undertake *lexical segmentation*, the perceptual process that enables a continuous stream of acoustic energy to be parsed into its individual word components (Jusczyk & Luce, 2002). Most recent accounts of lexical segmentation assume an integrative model in which listeners exploit a variety of perceptual strategies to successfully segment spoken language (McQueen & Cutler, 2001a). Based on the assumption that listeners will exploit the most economical means to achieve lexical segmentation, it is postulated that perceptual strategies may be dependent upon the quality of the acoustic signal and the richness of the contextual information (Mattys, et al., 2005). When segmental information affords insufficient cues, the MSS claims that listeners will utilise the presence of strong syllables to predict the onset of a new word (Cutler & Carter, 1987). Evidence of this strategy can be found in the patterns of LBEs made by listeners during attempts to decipher degraded speech (see Chapter 1, section 1.3.2.1 for full details of recognising connected speech).

The first phase of this research programme, Chapter 3, demonstrated that an experience involving either passive or explicit familiarisation with read passages produced by speakers with dysarthria facilitated immediate intelligibility improvements during subsequent transcription of dysarthric phrases. Furthermore, intelligibility gains were most pronounced under explicit conditions; that is, when the degraded stimuli was supplemented with written targets of the intended productions. Examination of possible sources of learning revealed that, despite discrepancies in intelligibility scores between conditions, both passive and explicit groups exhibited similar segmental benefits relative to the control group. However, the error patterns indicative of suprasegmental processing were remarkably different between the two condition groups. Specifically, the LBE patterns exhibited by listeners who received explicit familiarisation conformed strongly to MSS predictions, suggesting greater attention toward syllabic stress contrasts to inform word boundary decisions. In contrast, listeners who received passive familiarisation did not adhere to the predicted error patterns, reflecting a perceptual shift away from the anticipated prosodic perception cues. Thus, all listeners familiarised with dysarthric stimuli demonstrated improved intelligibility and attention toward acoustic-phonetic features, however the tendency to exploit syllabic stress cues for speech segmentation was only evident following explicit familiarisation. Accordingly, the study identified that learning mechanisms may differ depending on the familiarisation conditions (passive verses explicit) used to promote improved recognition of dysarthric speech.

Greenspan and colleagues (1988) reported that intelligibility scores from listeners familiarised with synthetic speech were influenced by the familiarisation stimuli; sentence-level stimuli was superior to word-level stimuli. Thus, it could be speculated that the source of learning may also be differentially influenced by the type of familiarisation stimuli used. However, previous studies with dysarthric speech have not observed a significant difference in intelligibility gains for listeners familiarised with word list versus paragraph-level stimuli (D'Innocenzo, et al., 2006; Tjaden & Liss, 1995a). It was noted that a systematic qualitative analysis of errors may be required to determine if any differential benefits are realised with respect to underlying cognitive-perceptual learning mechanisms (Tjaden & Liss, 1995a).

4.2.1 The Current Study

The current study further investigates the finding that the performance differences with passive or explicit conditions were more than simply magnitude of benefit. Its primary aim is to address why the listeners who received passive familiarisation in Chapter 3 exhibited differences in lexical segmentation compared to listeners who received explicit familiarisation. If the key variable is simply the presence or absence of written information, new listeners who receive either passive or explicit familiarisation with different types of familiarisation stimuli should elicit the same pattern as the prior results, with significant LBE pattern discrepancies. However, if the new familiarisation material specifically draws attention to the alternating syllabic stress of the transcription phrases, the condition discrepancy may disappear. This would provide evidence that the locus of learning during the familiarisation phase is dependent on the information (and density of that information) in the familiarisation procedure. Accordingly, the key research question asks: Do listeners familiarised with experimental phrases which emphasise syllabic stress cues improve their ability to exploit such cues regardless of whether learning conditions are passive or explicit? In addition, the study addressed whether there was an effect of familiarisation procedure, in which the magnitude of intelligibility and segmental gain was regulated by the type of familiarisation stimuli (passages versus phrases) and learning conditions (passive versus explicit).

4.3 METHOD

4.3.1 Research Design

A between-groups design was used to investigate perceptual learning effects associated with familiarisation procedures that varied with respect to stimuli (phrases versus passages) and condition (passive versus explicit). Two groups of listeners were familiarised with a set of 36 experimental stimuli produced by speakers with dysarthria under one of two experimental conditions: (1) auditory presentation of experimental phrases (passive-phrases), or (2) concurrent auditory and written presentation of experimental phrases (explicit-phrases). Following familiarisation, all listeners completed an identical phrase transcription test. Data from the current study was compared with the corresponding data from the two experimental groups in Chapter 3: (1) auditory presentation of passage readings (passive-passages), and (2) concurrent auditory and written presentation of passage readings (explicit-passages).

4.3.2 Listeners

Data were collected from 40 young healthy individuals (31 women, 9 men) with a mean age of 24.4 years (*SD* = 6.3). See Chapter 2, section 2.2 for further details of the listener participants. Data from 10 additional listeners (M = 25.8 years, *SD* = 4.7) who met the same inclusion criteria, was collected to provide a baseline measure of intelligibility of the testing speech set (section 4.3.3).

4.3.3 Speech Stimuli

Speech familiarisation and test material consisted of the 72 experimental phrases that comprised speech set one and speech set two (see Appendix B), re-labelled for use in the current study as *familiarisation speech set* and *test speech* set, respectively. See Chapter 2, section 2.5 for further details of the speech stimuli employed.

Baseline intelligibility of the test speech set was established by calculating the mean value of PWC scores from 10 listeners (see section 4.3.2) who transcribed the 36 phrases that made up the test speech set. These listeners received no prior familiarisation or training.[3] Transcription task instructions were identical to those used in the test phase in the perceptual learning procedure (as described in section 4.3.4). Based on this analysis, baseline intelligibility of the test speech set was established as 20.57% (*SD* = 3.4).

4.3.4 Perceptual Learning Procedure

The 40 listener participants were randomly assigned to one of two experimental groups, passive-phrases or explicit-phrases, so that each group consisted of 20 participants. The experiment was conducted in two distinct phases: (1) familiarisation phase, and (2) test phase. Figure 4.1 contains a diagrammatic representation of the perceptual learning procedure employed.

The experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 pro). Listeners were tested individually. The experiment was conducted via a laptop computer pre-loaded with the experimental procedure. Participants were told that they would undertake a listening task followed by a transcription task, and that task-specific instructions would be delivered via the computer program. This process was employed to ensure identical stimulus presentation methods across participants.

During the *familiarisation* phase, listeners in the passive-phrases experimental group were presented with auditory productions of the familiarisation speech set and were instructed to simply listen to the phrases. Listeners in the explicit-phrases experimental group were also presented with auditory productions of the familiarisation speech set, in addition to written transcripts of the intended phrase targets, and were instructed to read these alongside the auditory productions.

Immediately following the familiarisation task, both of the experimental groups participated in an identical *test* phase in which they transcribed the test speech set.

[3]The decision not to include a control group, in which listeners would be familiarised with phrases produced by neurologically intact speakers, was based on observations from Chapter 3. The findings of Chapter 3 indicated that listeners familiarised with dysarthric passages achieved significantly greater intelligibility benefits relative to a control group familiarised with neurologically intact speech. This observation would suggest that the inclusion of a control group is superfluous. Rather, a baseline intelligibility score of the testing material was established to provide a measure of control and validate any intelligibility gains realised with familiarisation.

Transcription task instructions were identical to that of the previous study. Phrases were presented one at a time and listeners were asked to listen carefully to each phrase and to type exactly what they heard. Listeners were told that all phrases contained real English words but that the phrases themselves would not make sense. They were told that some of the phrases would be difficult to understand, and that they should guess any words they did not recognise. Listeners were told to place an “X” to represent part of a phrase, if they were unable to make a guess. They were given 12 seconds to type each response. Presentation of all familiarisation and test phrase stimuli were presented in an entirely randomised manor for each of the 40 participating listeners.

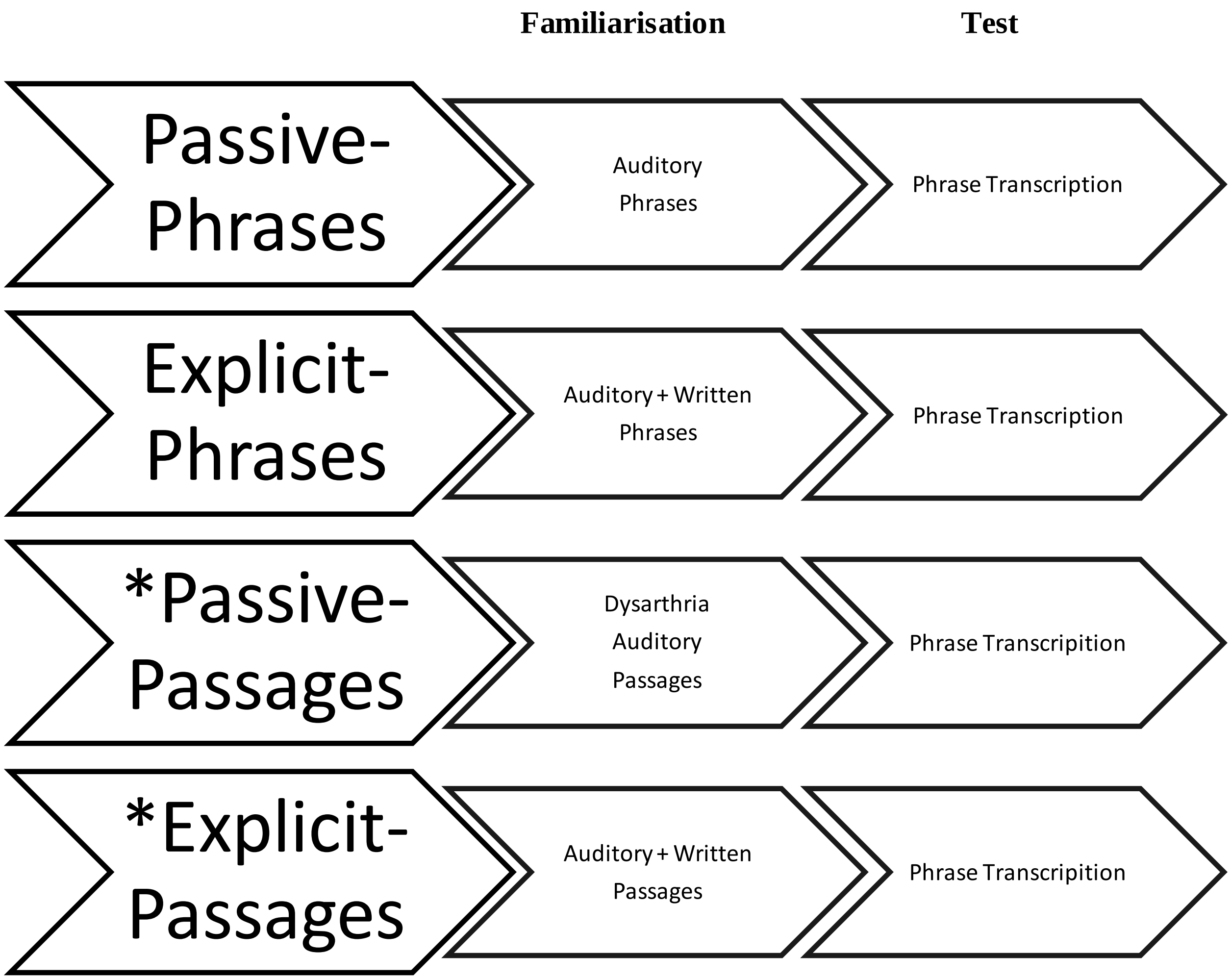

Figure 4.1. Perceptual learning procedure divided into two phases. The first column provides details of the familiarisation phase and the second provides details of the test phase. All stimuli were produced by speakers with dysarthria. * = data reported in Chapter 3.

4.3.5 Transcript Analysis

The total data set consisted of 40 transcripts of the 36 experimental phrases that made up the test speech set. The author independently analysed the listener transcripts for PWC, PSR, PSC, and the presence and type of LBEs. Details regarding the analysis and calculation of these measures are found in Chapter 2, section 2.6. The reliability of the transcript analysis for the 40 transcripts was then measured (details of the method of reliability measurement are reported in Chapter 2, section 2.6.4). The first author and a second trained judge reanalysed 25% of the transcripts. Discrepancies between the reanalysed data and the original data analysis are reported in terms of absolute mean difference. Pearson product-moment correlation coefficients were computed to assess the relationship between the data sets. Table 4.1 summarises the results. A strong, positive correlation between the reanalysed data and original data was found for analysis of the transcripts.

Table 4.1

Mean Difference and Pearson Product-Moment Correlation Coefficients for Intra- and Inter-judge Reliability of the Transcript Analysis

	Intra-judge		Inter-judge	
	MD (SD)	*r*	*MD (SD)*	*r*
PWC	0.43 (0.47)	.99*	0.92 (0.56)	.99*
PSR	0.70 (0.48)	.98*	2.10 (1.10)	.95*
LBE	0.50 (0.53)	.99*	1.60 (0.70)	.99*

Note. PWC = percent words correct; PSR = percent syllables correct; LBE = lexical boundary errors.

* $p < .001$

4.4 Results

4.4.1 Percent Words Correct

Figure 4.2 reflects the mean PWC scores for the two experimental groups familiarised with dysarthric phrases under passive (passive-phrases) or explicit (explicit-phrases) conditions. Figure 4.2 also includes corresponding data for listeners familiarised with dysarthric passages under passive (passive-passages) and explicit (explicit-passages) conditions from Chapter 3. Baseline intelligibility, as determined by a group of ten listeners who received no familiarisation, is included for comparative purposes.

A two-way ANOVA was conducted on PWC scores of the listeners familiarised with dysarthric speech, with condition (passive or explicit) and stimuli (passages or phrases) as between subject variables. The ANOVA revealed a significant main effect of condition, $F(1, 76) = 122.51$, $p < .001$, $\eta^2 = .27$. Thus, explicit familiarisation afforded significantly greater intelligibility gains than passive familiarisation. The main effect of stimuli was also significant, $F(1, 76) = 251.90$, $p < .001$, $\eta^2 = .55$. Thus, familiarisation with the passage stimuli afforded significantly greater intelligibility gains than familiarisation with the phrase stimuli. The interaction between condition and stimuli was not significant, $F(1, 76) = 70.58$, $p = .05$, $\eta^2 = .01$.

A one-way ANOVA was also conducted to compare the PWC scores of the listeners familiarised with passive-phrases and explicit-phrases with baseline intelligibility. The analysis revealed a significant difference in PWC scores across the three groups, $F(2, 47) = 67.17$, $p < .001$, $\eta^2 = .74$. *Post hoc* tests, using Bonferroni correction, demonstrated that the PWC scores achieved by the explicit-phrases group were significantly higher than both baseline intelligibility, $t(28) = 8.61$, $p < .001$, $d = 3.46$, and the passive-phrases group, $t(38) = 8.63$, $p < .001$, $d = 3.30$. However, there was no significant difference between PWC scores of the passive-phrases group and baseline intelligibility, $t(28) = 0.02$, $p = 1.0$, $d = .008$. Thus, intelligibility gains for listeners familiarised with the phrase stimuli were only realised under explicit conditions.[4]

[4] No significant difference in PWC scores for the passive-phrases and baseline intelligibility would suggest that the learning achieved by the explicit-phrases group can be attributed to the familiarisation procedure rather than learning something about the unique experimental stimuli.

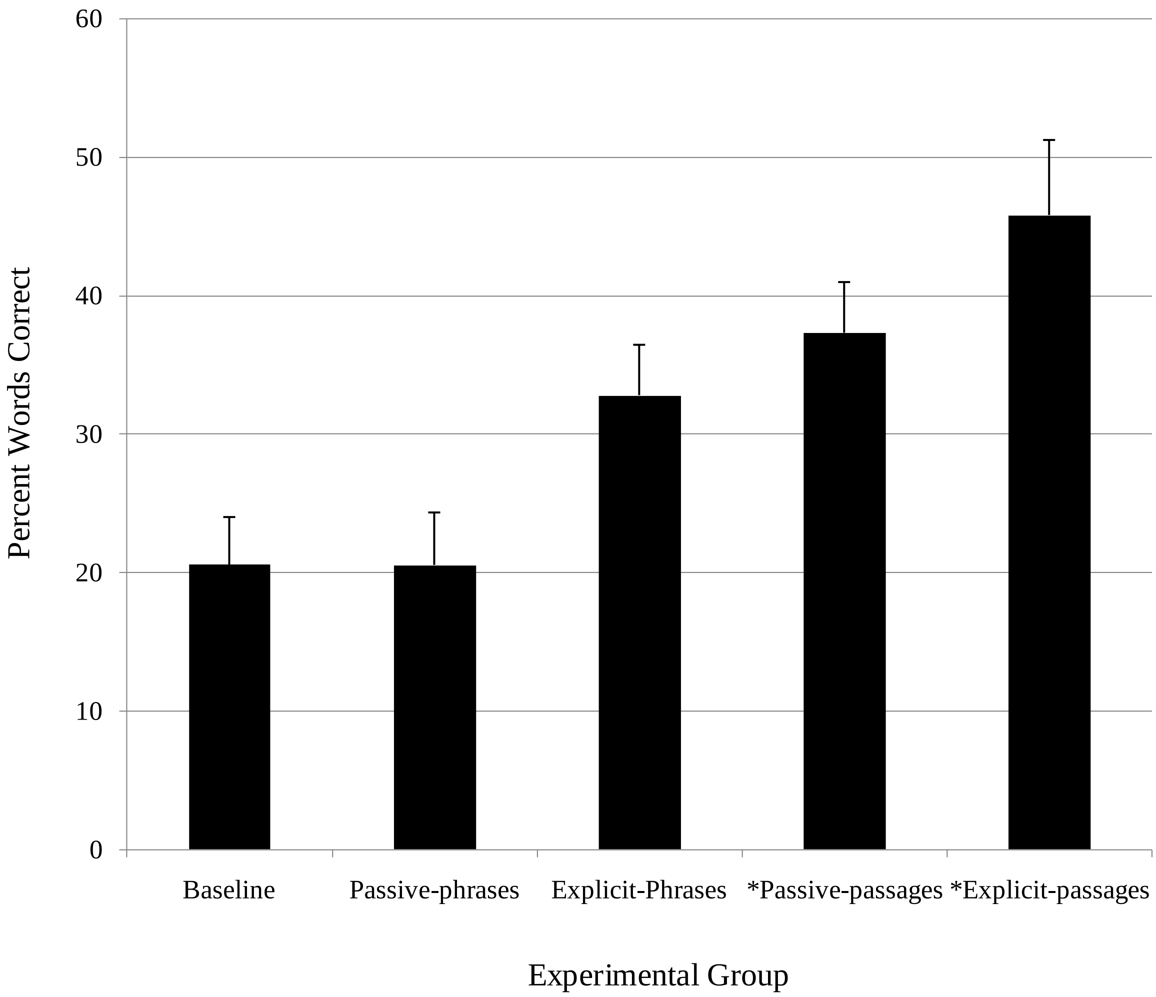

Figure 4.2. Mean percent words correct (PWC) for listeners by experimental group. Bars delineate + 1 standard deviation of the mean. * Data reported in Chapter 3.

4.4.2 Percent Syllable Resemblance

Figure 4.3 reflects the mean PSR scores, in addition to the mean PSC scores, for the two experimental groups familiarised with dysarthric phrases under passive (passive-phrases) and explicit (explicit-phrases) conditions. Figure 4.3 also includes corresponding data for listeners familiarised with dysarthric passages under passive (passive-passages) or explicit (explicit-passages) conditions. Pearson product-moment correlation coefficients demonstrated a strong relationship between the variables of PSC and PWC for both the passive-phrases and explicit-phrases groups (see Table 4.2). Accordingly, statistical analysis was performed on the PSR data only, as PSC findings are reflected in the analysis of PWC (see section 4.4.1).

A two-way ANOVA was conducted on PSR scores of the listeners familiarised with dysarthric speech, with condition (passive or explicit) and stimuli (passages or phrases) as between subject variables. The ANOVA revealed a small but significant main effect of condition, $F(1, 76) = 4.30$, $p = .04$, $\eta^2 = .05$. Thus, explicit familiarisation afforded greater benefits to a segmental measure of processing than passive familiarisation. There was no significant effect of stimuli, $F(1, 76) = 1.31$, $p = .26$, $\eta^2 = .02$, or an interaction effect between condition and stimuli, $F(1, 76) = 0.14$, $p = .71$, $\eta^2 = .002$.

Table 4.2

Mean Difference and Pearson Product-Moment Correlation Coefficients between Percent Words Correct and Percent Syllables Correct for Listeners by Experimental Group

Group [a]	*MD (SD)*	*r*
Passive-Phrases	7.1 (3.5)	.72*
Explicit-Phrases	6.3 (2.5)	.80*

[a] $n = 20$

* $p < .001$

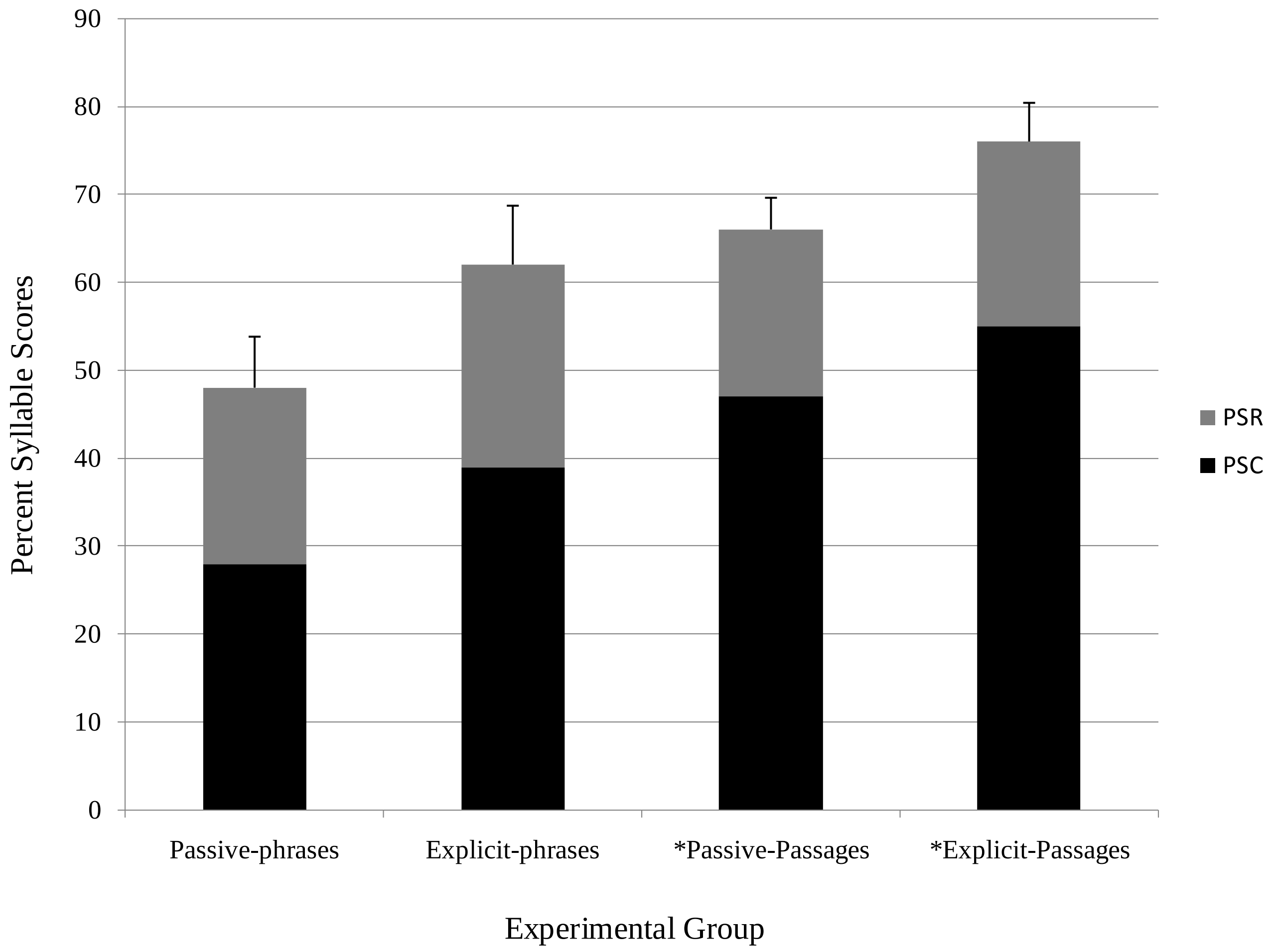

Figure 4.3. Mean percent syllable correct (PSC) and mean percent syllable resemblance (PSR) for listeners by experimental group. Bars delineate + 1 standard deviation of the mean PSR data. * = data reported in Chapter 3.

4.4.3 Lexical Boundary Error Patterns

Table 4.3 contains the LBE category proportions and the sum IS/IW and DW/DS ratios for the two experimental groups familiarised with dysarthric phrases under passive (passive-phrases) and explicit (explicit-phrases) conditions. Table 4.3 also includes corresponding data for listeners familiarised with dysarthric passages under passive (passive-passages) and explicit (explicit-passages) conditions. Contingency tables were constructed for the total number of LBEs by error type (i.e., insertion/deletion) and error location (i.e., before strong/weak syllable) for the passive-phrases and explicit-phrases groups to determine whether the variables were significantly related.

A within-group chi-square analysis revealed a significant interaction between the variables of type (insert/delete) and location (strong/weak) for the data generated by the passive-phrases group, $X^2(1, N = 20) = 71.84$, $p < .001$, and the explicit-phrases group, $X^2(1, N = 20) = 89.06$, $p < .001$. In both of these groups, erroneous lexical boundary insertions occurred more often before strong than before weak syllables, and erroneous lexical boundary deletions occurred more often before weak than before strong syllables. Such LBE error patterns are predicted (Cutler & Butterfield, 1992 see also Chapter 2, section 2.6.3). Ratio figures reflect the strength of adherence to these predicted error patterns—the greater the positive distance from “1,” the stronger the adherence. When compared to corresponding data from Chapter 3, the ratio values observed for the passive-phrases and explicit-phrases groups are similar to those observed for the explicit-passages group. The ratios values for both the passive-phrases and explicit-phrases groups support reliance on syllabic stress contrast cues to inform speech segmentation.

A between-group chi-square analysis was also used to examine differences in the distribution of errors exhibited by the passive-phrases and explicit-phrases groups. The comparison revealed no significant difference in error distribution between the two groups, $X^2(3, N = 40) = 3.9$, $p = .27$. Thus, the relative distribution of errors observed for the passive-phrases group were similar to those observed for listeners in the explicit-phrases group.

Table 4.3

Category Proportions of Lexical Boundary Errors Expressed in Percentages and Sum Error Ratio Values for Listeners by Experimental Group

Group [a]	%IS	%IW	%DS	%DW	IS/IW Ratio	DW/DS Ratio
Passive-phrases	51.60	16.80	11.40	20.20	3.1	1.8
Explicit-phrases	51.95	14.29	11.69	22.08	3.6	1.9
*Passive-passages	27.31	22.69	28.41	21.59	1.2	0.8
*Explicit-passages	42.42	12.31	16.70	28.57	3.5	1.7

Note: “IS”, “DS”, “IW” and “DW” refer to lexical boundary errors defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively.
* = data reported in Chapter 3 and included for visual comparison only.
[a] $n = 20$

4.5 DISCUSSION

The current investigation offers further insight into the learning mechanisms associated with improved recognition of dysarthric speech and provides evidence regarding the influence of the familiarisation procedure in this perceptual process. The primary finding was that listeners familiarised with experimental phrases exploited syllabic stress as a segmentation cue, regardless of whether the learning conditions were passive or explicit. This provides conclusive evidence that the key variable behind the use of particular segmentation strategies is not simply the presence or absence of written information during the familiarisation procedure. This finding, in conjunction with intelligibility scores and segmental processing data, is further discussed and elaborated on in the ensuing sections.

Key to an examination of underlying learning mechanisms is that intelligibility benefits for listeners familiarised with experimental phrases were only realised under explicit conditions; word recognition scores for listeners who received passive familiarisation to experimental phrases were no greater than baseline intelligibility of the testing phrases. This raises the question as to why listeners in the passive-phrases group did not achieve the moderate intelligibility benefit afforded by passive familiarisation with read passages. The most plausible explanation relates to the fundamental intelligibility differences between the phrase and passage familiarisation stimuli. Unlike the passages, which consisted of semantically and syntactically predictable sentences within a rich story context, the phrases were semantically anomalous. As such, listeners familiarised with the experimental phrases were disadvantaged in their capacity to deploy top-down, predictive processes to decipher the intended word targets. It was only when the lexical targets were provided (explicit-phrases) that the familiarisation stimuli were sufficiently intelligibility to enable listeners to extract information required for performance gains during subsequent encounters with the degraded speech.

Yet, despite significant performance differences, both passive and explicit groups familiarised with experimental phrases made LBE patterns that conformed to MSS predicted error patterns. That is, a greater number of predicted (IS and DW) versus non-predicted (IW and DS) errors (see Chapter 2, section 2.6.3 for detailed explanation). Lexical boundary error patterns did not reflect the condition discrepancy observed in the initial study and thus, the speculation that the learning mechanisms vary depending on the presence of absence of written information, is not supported by the current data. Adherence to predicted error patterns denotes a reliance on strong syllables to identify word onsets and ratio figures reflect the strength of adherence to such patterns (Cutler & Carter, 1987). Both groups familiarised with experimental phrases elicited strong adherence to the predicted patterns, similar to the level of adherence observed for the explicit-passages group in Chapter 3 (see Table 4.2). This provides compelling evidence that the experimental phrases served to direct attention to the cue of syllabic stress for making lexical boundary decisions. What is interesting, however, is that these suprasegmental changes were evident even in the absence of improved intelligibility performance (passive-phrases group). This may indicate that learning to better detect syllables stress contrast cues precedes the realisation of any significant intelligibility improvements. In the initial study, which also examined stability of perceptual learning over time, the possibility that learning decay may occur at different rates across different levels of analysis was raised (see Chapter 3, section 3.5.2). The current LBE data, in conjunction with PWC data, would suggest the same may be true for *learning* across different levels of analysis. It appears that changes to suprasegmental processing may emerge faster than evidence of intelligibility benefits.

Analysis of errors indicative of segmental processing found a small but significant effect of condition; listeners familiarised with dysarthric speech under explicit conditions achieved a significantly higher percentage of syllables that bore phonemic resemblance to the targets relative to listeners familiarised with dysarthric speech under passive conditions. This analysis aimed to tap the extent to which the familiarisation process promoted acoustic-phonetic mapping that could be subsequently leveraged by the listeners. Thus, it appears that the provision of written information regarding the lexical targets may have aided perceptual mapping of acoustic information onto existing mental representations.

Comparisons of intelligibility data revealed that performance benefits were regulated by both stimuli and learning conditions. Listeners familiarised with the passage-level stimuli performed significantly better than listeners familiarised with the experimental phrases. This finding seems particularly robust given the transfer-appropriate processing theory which postulates that improvements may be magnified when learning conditions are reinstated at testing (e.g., Lockhart, 2002; Rajaram, Srinivas, & Roediger, 1998). Unlike the experimental phrases, the read passages did not afford similarities to the test stimuli. That passage-level familiarisation stimuli afforded superior intelligibility gains is consistent with early work with synthetic speech (Greenspan, et al., 1988), however, more recent studies with noise-vocoded speech suggest otherwise (Davis, et al., 2005; Loebach, et al., 2010). Davis and colleagues (2005) reported significantly greater word recognition scores for listeners familiarised with real word versus nonword sentences, but observed no performance difference for listeners familiarised with semantically meaningful versus syntactic prose sentence stimuli. It was concluded that lexical information may inform perceptual learning of sentence-level stimuli, but that this learning can transpire in the absence of sentence-level meaning (see Chapter 1, section 1.4.1.1). Findings were replicated in a recent study by Loebach, Pisoni, and Svirsky (2010). While existing studies with dysarthric speech have also observed no difference in word recognition scores for listeners familiarised with word list versus paragraph stimuli (D'Innocenzo, et al., 2006; Tjaden & Liss, 1995a), Tjaden and Liss found a pervasive trend of improved performance for listeners familiarised with paragraph stimuli produced by a speaker with spastic-ataxic dysarthria of 46% single word intelligibility. It was speculated that the disparity in type and severity of the dysarthric speech stimuli may explain the absence of this trend in the later study, wherein listeners were familiarised with stimuli produced by a speaker with flaccid-spastic dysarthria of 60% single word intelligibility (D'Innocenzo, et al., 2006). These findings, in conjunction with the current data comparisons, suggest that reliance on linguistically-informative familiarisation stimuli to support perceptual learning of dysarthric speech may increase as intelligibility levels decrease. This speculation may similarly be applied to the evidence that listeners familiarised with dysarthric speech under explicit conditions outperformed listeners familiarised with dysarthric speech under passive conditions.

Taken together, the intelligibility data offer preliminary evidence that signal-independent information afforded by the stimuli and/or learning conditions may promote perceptual learning of this type and severity of signal degradation. Certainly there is an abundance of evidence that signal-independent information can improve perception of dysarthric speech (e.g., Dongilli, 1994; Garcia & Cannito, 1996; Garcia & Dagenais, 1998; Hammen, Yorkston, & Dowden, 1991; Hustad & Beukelman, 2001; Vogel & Miller, 1991; Yorkston, Dowden, & Beukelman, 1992). For example, Hustad and Beukelman (2001) showed increased intelligibility when severely dysarthric speech was supplemented with linguistic information in the form of alphabet cues and/or topic cues. Similarly, intelligibility of the dysarthric signal was reported greater in the context of highly predictable sentences when compared with sentences with low inter-word predictability (Garcia & Cannito, 1996; Garcia & Dagenais, 1998), and greater in the context of a sentence when compared with recognising single words in isolation (Dongilli, 1994; Yorkston & Beukelman, 1978). These findings are comparable with Lindblom's (1990) model of mutuality which postulates that when signal information is poor, signal-independent information can be used to enhance understanding of the degraded input. It appears that information independent of the acoustic signal may act as an external cue to enhance perception of dysarthric speech. The current finding suggests that signal-independent information may also enhance perceptual learning of moderate hypokinetic dysarthria.

Future research that investigates the role of signal-independent information in perceptual learning across different types and severities of dysarthria is required to apply findings more generally to perceptual learning of the neurologically degraded speech signal. Furthermore, as the current study was not designed to serve as a test of efficacy regarding different familiarisation stimuli, a number of factors were not controlled for. Listeners familiarised with read passages heard a total of 57 sentences (3 passages [19 sentences] x 3 speakers), whereas listeners familiarised with experimental phases heard a set of 36 phrases

(12 phrases x 3 speakers). Thus, a greater amount of exposure for listeners familiarised with the read passages may have interfered with the perceptual learning outcomes. In addition, no attempt was made to balance word familiarity and word frequency in the passage and phrase stimuli, although both familiarity and frequency have been identified as important factors in recognising spoken language under degraded conditions (Howes, 1957; G. A. Miller, Heise, & Lichten, 1951). Studies that control for the amount and content of the familiarisation stimuli are required to strengthen the present findings.

4.6 CONCLUSION

This study has provided further evidence that performance discrepancies cannot be fully explained in terms of magnitude of benefit. Findings revealed that the key variable behind the use of particular segmentation strategies was not simply the presence or absence of written information during familiarisation. Rather, it appears that the ability to exploit syllabic stress contrasts cues for lexical boundary decisions necessitates some level of prompting—whether that is written cues afforded by explicit conditions or experimental stimuli that emphasises prosodic patterns. Thus, while intelligibility gains were superior when learning conditions were explicit (relative to passive) or when passage-level stimuli (relative to experimental phrases) was employed, underlying error patterns would suggest that that the locus of learning is influenced differentially by the information afforded within the familiarisation procedure.

CHAPTER FIVE

Phase Three: The Role of Linguistic and Indexical Information in Improved Recognition of Dysarthric Speech

5.1 ABSTRACT

Chapters 3 and 4 demonstrated that the intelligibility benefits afforded by experience with dysarthric speech were realised only when the linguistic properties of the signal were emphasised with signal-independent information. However, the speech signal carries both linguistic and indexical (speaker-specific) properties. It is currently not known how the indexical features of dysarthric speech influence linguistic processing of the signal. This investigation forms the final phase of the current research programme and investigates the role of indexical information in perceptual learning of dysarthric speech. Forty listeners were randomly assigned to one of two identification training tasks, aimed at highlighting either the linguistic or indexical properties of the dysarthric signal. Immediately following familiarisation, all listeners completed an identical phrase transcription task. Analysis of post-training listener transcripts revealed remarkably similar intelligibility improvements for listeners trained to attend to either the linguistic (word identification task) or the indexical (speaker identification) properties of the signal. Perceptual learning effects were also evaluated with regards to underlying error patterns indicative of segmental and suprasegmental processing. Comparisons revealed no significant difference at either level of perceptual processing for the two training groups. The findings of this study suggest that elements within both the linguistic and indexical properties of the dysarthric signal are learnable and interact to promote improved processing of this type and severity of speech degradation. Furthermore, error pattern analysis indicates that similar cognitive-perceptual mechanisms may underlie the processing of indexical and linguistic information. Thus, the current study extends support for the development of a model of perceptual processing in which the learning of indexical properties is encoded and retained alongside linguistic properties of the signal (e.g., Pisoni, 1997).

5.2 INTRODUCTION

Speech perception involves extracting relevant information from both *linguistic* and *indexical* properties within the signal. Linguistic information conveys the content of the utterance. This includes phonological, morphological, syntactic, and semantic information afforded within the word, phrase, and sentence structures of the acoustic signal (Levi & Pisoni, 2007). Indexical information, however, refers to any of the extralingustic elements within the signal that index specific speaker attributes (Abercrombie, 1967). This includes information pertaining to the speaker's gender (e.g., Munson, et al., 2006), regional dialect (e.g., Hagiwara, 1997; Hillenbrand, et al., 1995), or emotional state (e.g., Costanzo, et al., 1989; Murry & Arnott, 1993) (see Chapter 1, section 1.3.2 for a full discussion).

Chapters 3 and Chapter 4 of the current research programme demonstrated that experience-evoked intelligibility improvements with processing of dysarthric speech were only realised when the familiarisation procedure emphasised linguistic properties of the dysarthric signal—with linguistically-predictive passage level stimuli and/or supplementary written information. Based on this, it is plausible to assume a linguistic influence in improved recognition of dysarthric speech. However, research has yet to identify if indexical properties of the dysarthric signal can also promote perceptual learning of the neurologically degraded speech signal.

Founded on the premise that the perceptual system disregards any speaker-specific variation in an attempt to normalise the signal to a stable linguistic form, conventional models of spoken language processing have largely ignored indexical properties of the signal (e.g., Halle, 1985, see Chapter 1, section 1.3.2 for full details). However, these traditional perception paradigms are challenged by the growing body of work that has observed a speaker-specific influence on perceptual processing (again, see Chapter 1, section 1.3.2). In brief, research has identified a perceptual benefit of indexical consistency, documenting improved signal processing under single- versus multiple-speaker stimulus presentation conditions (e.g., Creelman, 1957; Goldinger, et al., 1991; Mullennix, et al., 1989). This supports the idea that listeners encode and retain indexical elements of the acoustic signal, alongside processing of the linguistic information (e.g., Pisoni, 1997). Furthermore, there is preliminary evidence to suggest that familiarisation with indexical properties of the signal may also facilitate intelligibility benefits when processing artificially modified speech signals (see Chapter 1, section 1.4.1.5 for full study details). To summarise, improved recognition of speech in noise was observed when novel words and sentences were produced by familiar (listeners received prior training to identify the speakers by name) versus unfamiliar speakers (Nygaard & Pisoni, 1998; Nygaard, et al., 1994). Similarly, perceptual benefits of prior training to attend to speaker-specific signal properties were observed with improved recognition of noise-vocoded speech (Loebach, et al., 2008). Nygaard and Pisoni (1998) have postulated that that encoding of both linguistic and indexical properties may recruit similar cognitive-perceptual processes.

Research has yet to document the exact learning mechanisms associated with indexical processing, but one hypothesis is that listeners extract something from the indexical regularities of these systematically degraded speech signals that enables improved perceptual processing during subsequent encounters. If indexical properties provide a source of learning for processing of speech in noise or noise-vocoded speech, one may readily assume the same to be true for all forms of speech degradation. However, a significant challenge arises when attempting to adopt phenomena observed in experiments using highly constrained artificially degraded speech to that of the neurologically degraded speech (see Chapter 1, section 1.4.2). To illustrate, noise-vocoded speech is created by the systematic removal of specific spectral aspects of the acoustic signal (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). However, dysarthric speech is produced upon a platform of impaired muscle tone, inadequate respiratory support, phonatory instability, and deficient articulatory movement. The implication for speech production is that while some acoustic degradation may be relatively consistent, other breakdowns in speech occur in nonsystematic and unpredictable ways speech (see Chapter 1, section 1.4.2). Thus, the influence of indexical information in perceptual learning of dysarthric speech is currently unknown.

An understanding of the role that indexical information plays in perceptual learning of dysarthric speech is imperative to establish a theoretical framework that supports the development of listener-based treatment for the management of this neurologically speech disorder. Such knowledge may also have implications for current models of perceptual processing. Mattys and Liss (2008) have reported on the perceptual benefit of indexical consistency for the processing of dysarthric words presented by the same, versus a different speaker (see Chapter 1, section 1.3.4); however, research has yet to investigate whether training to attend to indexical properties of the dysarthric signal can facilitate improved recognition of this type of speech degradation. Furthermore, evidence of the associated cognitive-perceptual changes that may transpire with learning of indexical properties has yet to be documented.

5.2.1 The Current Study

The purpose of the current study was to investigate whether directing attention towards indexical information within the dysarthric signal could facilitate improved perceptual learning of this type of speech and also how this learning compares to that afforded by attention towards the linguistic properties of the signal. The present study addressed the following key questions: (1) Do listeners trained to attend to the indexical properties of the dysarthric signal demonstrate similar intelligibility benefits as those achieved by listeners trained to attend to the linguistic information; and (2) Does training to attend to indexical versus linguistic properties differentially influence error patterns at segmental and suprasegmental levels of perceptual processing?

5.3 METHOD

5.3.1 Research Design

A between-group design was used to investigate perceptual learning effects for listeners familiarised with dysarthric speech via one of two types of training: (1) linguistic training (word identification task), or (2) indexical training (speaker identification task). Following training, all listeners engaged in an identical transcription task with 36 novel phrases produced by the speakers with dysarthria.

5.3.2 Listeners

Primary data were collected from 40 young healthy individuals (29 women, 11 men) with a mean age of 24.1 years (*SD* = 6.3). See Chapter 2, section 2.2 for further details of the listener participants.

5.3.3 Speech Stimuli

Familiarisation material consisted of readings of the Rainbow Passage (Fairbanks, 1960) (see Appendix A) by speakers with dysarthria. Training and test stimuli consisted of the 72 experimental phrases that made up speech set one and speech set two (see Appendix B), re-labelled for use in the current study as *training speech set* and *test speech set*, respectively. See Chapter 2, section 2.5 for further details of the speech stimuli employed.

Baseline intelligibility of the test speech set was ascertained in Chapter 4, where the mean value of the PWC scores from ten listeners who transcribed the test speech was calculated. These listeners received no prior familiarisation or training (see Chapter 4, section 4.3, for further details). Based on this analysis, baseline intelligibility of the test speech set was established as 20.57% (*SD* = 3.4).

5.3.4 Perceptual Learning Procedure

The 40 listener participants were randomly assigned to one of two training conditions, *word identification* (linguistic) or *speaker identification* (indexical)[5], so that each experimental group consisted of 20 participants. The experiment was conducted in three distinct phases: (1) familiarisation phase, (2) training phase, and (3) test phase. Figure 5.1 contains a diagrammatic representation of the perceptual learning procedure employed.

[5]The decision not to include control groups, in which listeners would receive linguistic or indexical training on neurologically intact speech, was based on observations from Chapter 3. The findings of Chapter 3 indicated that listeners familiarised with dysarthric passages achieved significantly greater intelligibility scores relative to a control group familiarised with neurologically intact passages. This observation would suggest that inclusion of control training groups is superfluous. Rather, the baseline intelligibility score of the test material (established in Chapter 4) was used to provide a measure of control and validate any intelligibility gains realised with training.

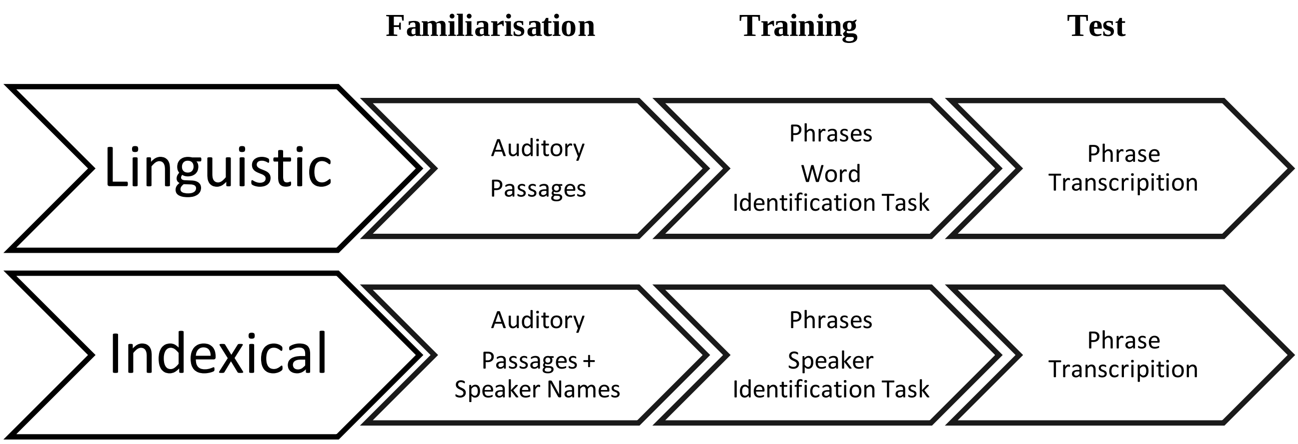

*Figure 5.1.*Perceptual learning procedure divided into three phases. The first column provides details of the familiarisation phase, the second provides details of the training phase, and the third provides details of the test phase. All stimuli were produced by speakers with dysarthria.

The experiment was conducted in a quiet room using sound-attenuating headphones (Sennheiser HD 280 pro). Listeners were tested individually. The experiment was presented via a laptop computer pre-loaded with the experimental procedure. Participants were told that they would undertake a listening task followed by a transcription task, and that task-specific instructions would be delivered via the computer programme. This process was employed to ensure identical stimulus presentation methods across participants.

During the *familiarisation* phase, listeners in both experimental groups were presented with three readings of the rainbow passage, each produced by a different speaker with dysarthria. To ensure each speaker was heard in each position a similar number of times, the order in which each of the 20 participants in each experimental group heard the three speakers was counterbalanced. For example, two of the speakers were heard in the first position seven times and one speaker six times, with similar ratios for the second and third positions. The order was then randomized using the Knuth implementation of the Fisher-Yates shuffling algorithm (Knuth, 1998). In addition to the readings, listeners in the speaker identification group also received the name[6] of the speaker producing the passage (John, Bob, or Peter). All listeners were informed of the nature of their subsequent task and given relevant instructions regarding attention allocation during familiarisation with the passage readings—

[6] Names changed to comply with participant confidentiality agreement.

listeners in the word identification group were instructed to listen carefully to any information that may help them learn to recognise what was being said and listeners in the speaker identification group were instructed to listen carefully to any information that may help them learn to recognise the speaker.

Immediately following the familiarisation phase, participants engaged in the *training* phase, which involved the 36 experimental phrases that made up the training speech set. Following the presentation of each individual phrase, listeners engaged in either a word or speaker identification task. Listeners in the word identification group were presented with three words and asked to use the mouse to select which word they thought they heard within the phrase. They were told that they would have heard only one of the three words. Listeners were given as long as required to make their word selection. Upon selection of a word choice, regardless of accuracy, the correct response was highlighted as feedback regarding task performance. Listeners in the speaker identification task were presented with the names of all three speakers and asked to use the mouse to select the speaker they thought they heard. As with the listeners in the word identification group, these listeners were given as long as required to make their name selection, and upon their selection of a name, the correct response was highlighted. The training phrases were presented randomly to each of the 40 listeners.

Immediately following the training task phase, both experimental groups participated in an identical *test* phase in which they transcribed the 36 novel phrases that made up the test speech set. Transcription task instructions were identical to those of the previous two studies. Phrases were presented one at a time and listeners were asked to listen carefully to each phrase and to type exactly what they heard. Listeners were told that all phrases contained real English words but that the phrases themselves would not make sense. They were told that some of the phrases would be difficult to understand, and that they should guess any words they did not recognise. Listeners were told to place an “X” to represent part of a phrase, if they were unable to make a guess. They were given 12 seconds to type each response. The 36 phrases that made up the test speech set were presented randomly to each of the 40 listeners.

In order to ensure listeners trained with either the word or speaker identification task recognised the desired properties within the signal, linguistic or indexical respectively, a 70% criterion[7] across the 36 training items was selected. The software program that delivered the perceptual learning paradigm automatically identified whether a response was “correct” or “incorrect” on the word or speaker identification task. Responses were then tallied across the 36 items and converted into a single percent item correct score for each individual listener. Figure 5.2 shows that listeners all performed above the 70% criterion on the training task and subsequently, the final analysis involved analysis of all 20 listener transcripts per training group (see section 5.3.5). An independent *t*-test between percent correct identification for listeners who received the word identification training task (M = 77.36, SD = 4.2) and listeners who received the speaker identification training task (M = 77.56, SD = 4.6) revealed no statistically significant difference between the two training groups, $t(38) = 0.08$, $p = .97$, $d = .02$. This would suggest that similar levels of attention towards the intended training targets across the two experimental groups was achieved.

[7] Based on the study by Nygaard and Pisoni (1998) in which the authors employed a 70% criterion to separate “good” from “poor” learners.

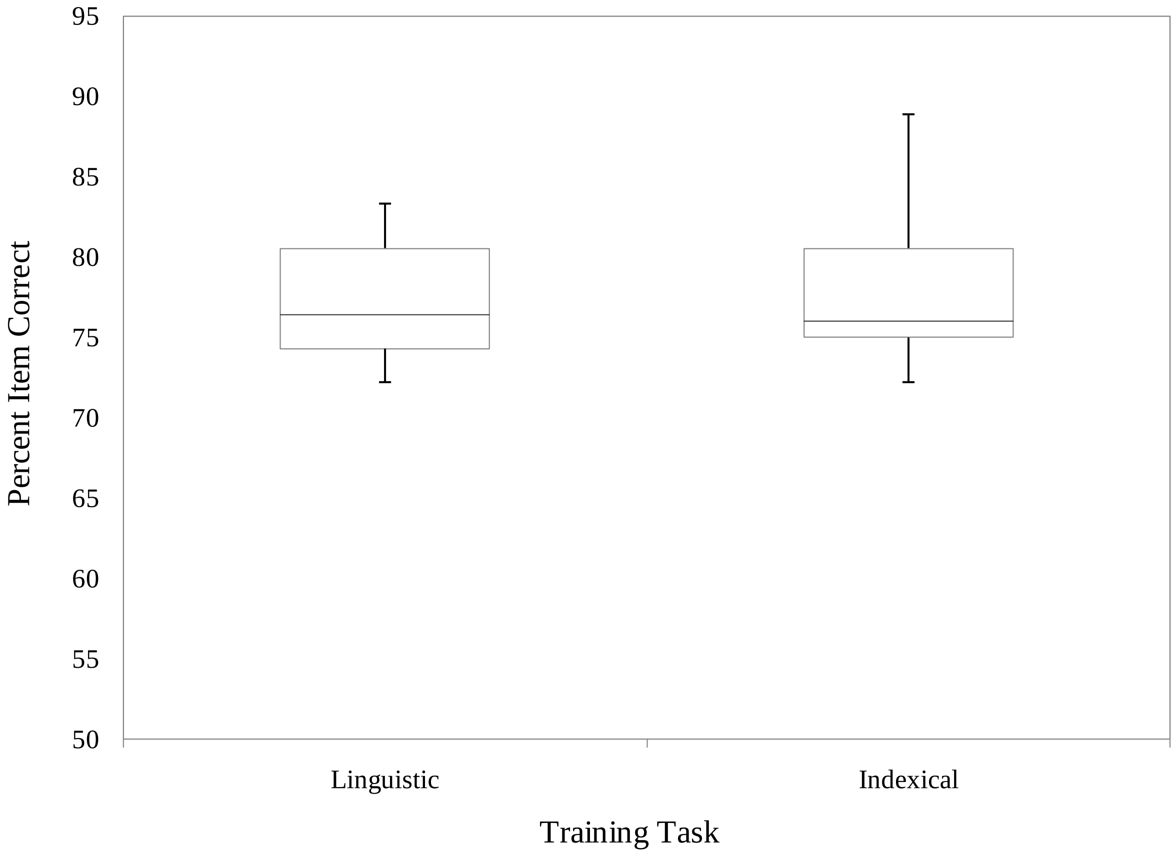

Figure 5.2. Percent items correct for listeners trained to attend to linguistic information (*n* = 20) and for listeners trained to attend to indexical information (*n* = 20).

5.3.5 Transcript Analysis

The total data set consisted of 40 transcripts of the 36 experimental phrases that made up the test speech set. The author independently analysed the listener transcripts for PWC, PSR, PSC, and the presence and type of LBEs. Details regarding the analysis and calculation of these measures are found in Chapter 2, section 2.6. The reliability of the transcript analysis for the 40 transcripts was then measured (details of the method of reliability measurement are reported in Chapter 2, section 2.6.4). The first author and a second trained judge reanalysed 25% of the transcripts. Discrepancies between the reanalysed data and the original data analysis are reported in terms of absolute mean difference. Pearson product-moment correlation coefficients were computed to assess the relationship between the data sets. Table

5.1 summarises the results. A strong, positive correlation between the reanalysed data and original data was found for analysis of the transcripts.

Table 5.1

Mean Difference and Pearson Product-Moment Correlation Coefficients for Intra- and Inter-judge Reliability of the Transcript Analysis

	Intra-judge		Inter-judge	
	MD (SD)	*R*	*MD (SD)*	*R*
PWC	0.21 (0.34)	.99*	0.52 (0.46)	.99*
PSR	0.40 (0.52)	.97*	1.00 (0.47)	.91*
LBE	0.30 (0.48)	.99*	0.90 (0.57)	.98*

Note. PWC = percent words correct; PSR = percent syllables correct; LBE = lexical boundary errors.

* $p < .001$.

5.4 RESULTS

5.4.1 Percent Words Correct

Figure 5.3 reflects the mean PWC scores for listeners familiarised with dysarthric speech via either a linguistic or indexical training task. Baseline intelligibility, as determined by a group of ten listeners who received no prior training (see section 5.3.3), is included for comparative purposes. A one-way ANOVA showed a significant effect of group for PWC scores following familiarisation, $F(2, 47) = 13.9$, $p < .001$, $\eta^2 = .37$. *Post hoc* tests, using Bonferroni correction, indicated that PWC scores of listeners in both the indexical, $t(28) = 4.84$, $p < .001$, $d = 1.85$, and linguistic, $t(28) = 4.81$, $p < .001$, $d = 2.17$, training groups were significantly higher than the baseline intelligibility. There was no significant difference in PWC scores between the linguistic or indexical training groups, $t(38) = 0.04$, $p = .999$, $d =$

.01. Thus, similar intelligibility gains were observed for the listeners who received linguistic training and the listeners who received indexical training.

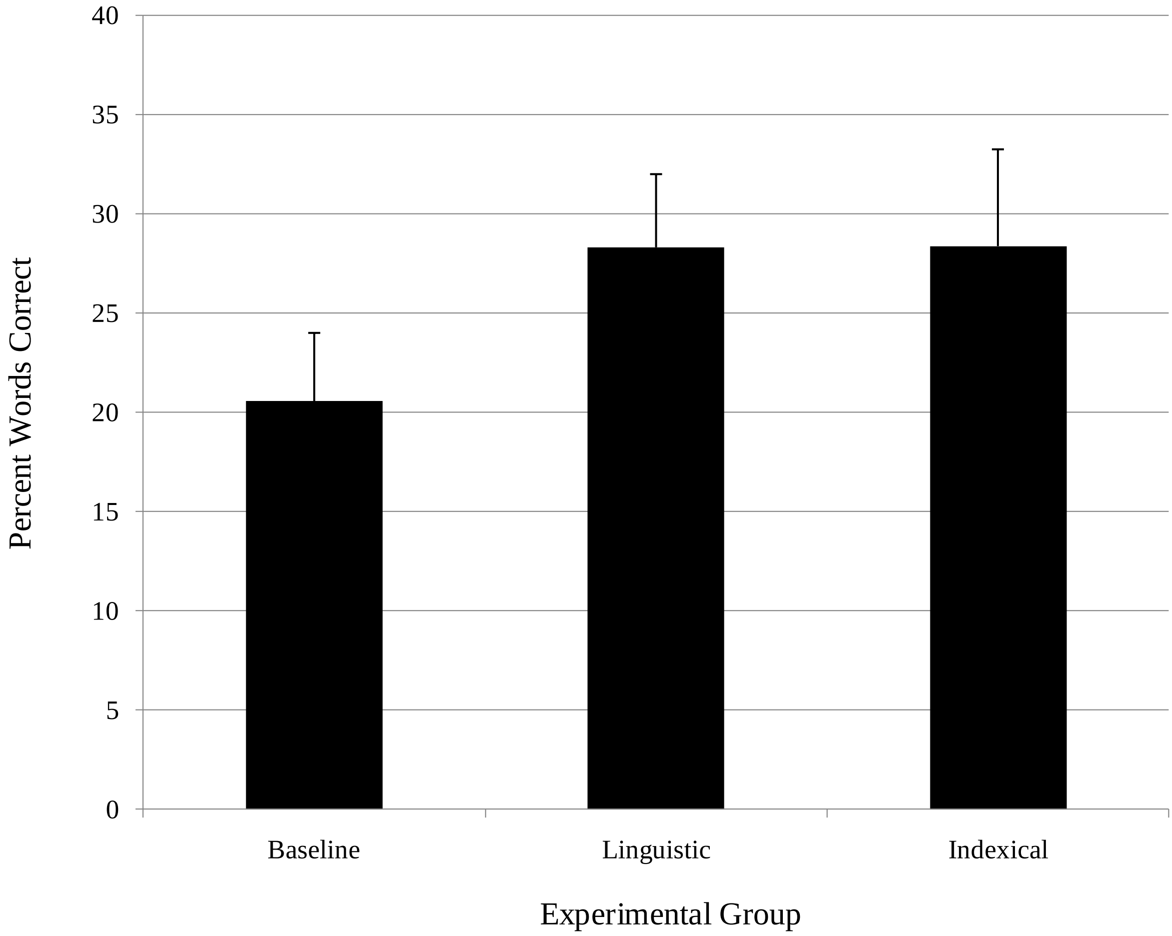

Figure 5.3. Mean percent words correct (PWC) for listeners by experimental group. Bars delineate + 1 standard deviation of the mean.

5.4.2 Percent Syllable Resemblance

Figure 5.4 displays the mean PSR scores, in addition to the mean PSC scores, for listeners familiarized dysarthria speech via either a linguistic or indexical training task. Pearson product-moment correlation coefficients demonstrated a strong relationship between the variables of PSC and PWC for both training groups (see Table 5.2). Accordingly, statistical analysis was performed on the PSR data only, as PSC findings are reflected in the analysis of PWC (see section 5.4.1). An independent groups *t*-test revealed no significant difference in PSR scores achieved by the two training groups, $t(38) = 1.01$, $p = 3.20$, $d = .32$. Thus, the indexical training task facilitated similar reliance on a segmental measure of perceptual processing as the linguistic training task.

Table 5.2

Mean Difference and Pearson Product-Moment Correlation Coefficients between Percent Words Correct and Percent Syllables Correct for Listeners by Experimental Group

Group [a]	*MD (SD)*	*r*
Linguistic	7.11 (2.69)	.75*
Indexical	7.97 (3.23)	.84*

[a] $n = 20$

* $p < .001$

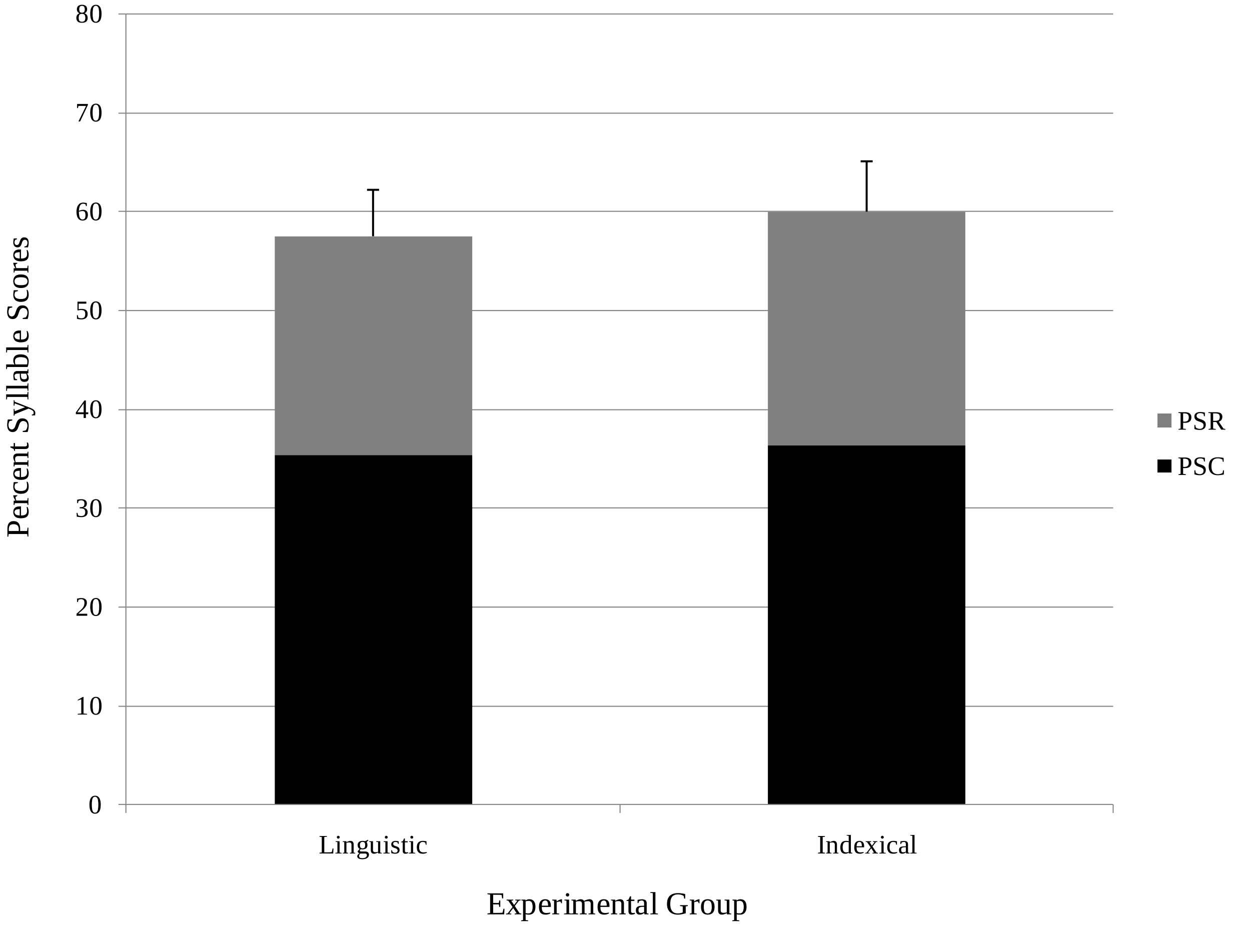

Figure 5.4. Mean percent syllable correct (PSC) and mean percent syllable resemblance (PSR) for listeners by experimental group. Bars delineate + 1 standard deviation of the mean PSR data.

5.4.3 Lexical Boundary Errors

Table 5.3 contains the LBE category proportions and the sum IS/IW and DW/DS ratios for listeners familiarised with dysarthric speech via either a linguistic or indexical training task. Contingency tables were constructed for the total number of LBEs by error type (i.e., insertion/deletion) and error location (i.e., before strong/weak syllable) for the two training groups to determine whether the variables were significantly related. A within-group chi-square analysis revealed a significant interaction effect between the variables of type (insert/delete) and location (strong/weak) for the data generated by the group of listeners who received the linguistic training task, $X^2(1, N = 20) = 47.57$, $p < .001$, and by the group of listeners who received the indexical training task, $X^2(1, N = 20) = 73.10$, $p < .001$. Thus, in both of the training groups, erroneous lexical boundary insertions occurred more often before strong than before weak syllables, and erroneous lexical boundary deletions occurred more often before weak than before strong syllables. Such LBE error patterns are predicted (Cutler & Butterfield, 1992 see also Chapter 2, section 2.6.3). Ratio figures reflect the strength of adherence to these predicted error patterns—the greater the positive distance from “1,” the stronger the adherence. Thus, ratio values for both the linguistic and indexical training groups support a strong reliance on syllabic stress contrast cues to inform speech segmentation. However, the magnitude of the IS/IW ratio is substantially greater for the indexical training group relative to the linguistic training group. Thus, the ratio values also indicate that reliance on syllabic stress contrast cues may be greater for the listeners who received indexical training.

A between-group chi-square analysis was used to examine differences in the distribution of errors exhibited by the linguistic and indexical training groups. The comparison revealed no significant difference in error distribution between the two groups, $X^2(3, N = 40) = 4.50$, $p = .21$. Thus, the relative distribution of errors observed for the linguistic training group were similar to those observed for the indexical training group.

Table 5.3

Category Proportions of Lexical Boundary Errors Expressed in Percentages and Sum Error Ratio Values for Listeners by Experimental Group

Group [a]	%IS	%IW	%DS	%DW	IS/IW Ratio	DW/DS Ratio
Linguistic	51.72	19.27	11.56	17.44	2.7	1.5
Indexical	54.53	16.21	10.53	18.74	3.4	1.8

Note: “IS”, “DS”, “IW” and “DW” refer to lexical boundary errors defined as insert boundary before strong syllable, delete boundary before strong syllable, insert boundary before weak syllable, and delete boundary before weak syllable, respectively.

[a] $n = 20$

5.5 DISCUSSION

The two earlier investigations of this research programme established the perceptual benefit of a familiarisation procedure in which linguistic properties of the dysarthric stimuli were emphasised. Accordingly, the present investigation sought to identify if perceptual learning of dysarthric speech is also influenced by indexical information within the signal. The current study observed intelligibility improvements following training on indexical properties of the signal, and moreover, that these paralleled the performance gains achieved following training on linguistic signal properties. In addition, error patterns at both segmental and suprasegmental levels of perceptual processing for listeners who received indexical training were remarkably similar to those exhibited by listeners who received linguistic training. Thus, the perceptual learning afforded by an indexical training task was comparable to that which occurred following a linguistic training task. Findings and implications are further discussed with regards to models of perceptual processing.

Listeners who completed a training task that emphasised the indexical properties of the neurologically degraded signal achieved intelligibility scores that were significantly higher than the baseline intelligibility of the data set. Thus, it appears that attention to the indexical elements of the dysarthric signal may provide a source of learning in perceptual adaption to this type and severity of speech degradation. While different perceptual learning paradigms were employed, the findings validate those reported by Nygaard and colleagues (1998; 1994) wherein improved linguistic processing of speech in noise was observed with prior training to identify the names of the 10 speakers providing the speech stimuli. Furthermore, the present study found that intelligibility improvements following an indexical training task were virtually identical to those observed for listeners familiarised with a training task in which linguistic properties were highlighted. Comparable intelligibility scores, regardless of training type, demonstrated that directing perceptual attention towards indexical elements of the signal afforded a similar performance gain to that achieved by directing attention towards the linguistic properties. This finding is consistent with previous research using noise-vocoded speech which found that intelligibility scores following familiarisation with indexical elements of the signal (speaker identification task) were equivalent to those following familiarisation with linguistic elements of the signal (transcription task) (Loebach, et al., 2008). Thus, the current findings reveal that the perceptual benefit of indexical information on processing of speech that has been systematically degraded continues to be robust under the highly variable and frequently inconsistent acoustic degradation that is associated with the dysarthric signal.

From the performance data alone, the following two conclusions can be drawn: that training to attend to indexical properties of the neurologically degraded signal does provide some perceptual benefit (relative to no training), and that this level of benefit is similar to that which is afforded by training on the linguistic aspects of the signal. Traditional views of perceptual processing do not account for the processing of speaker-specific detail, and thus the current findings extend support for the development of new theoretical paradigms in which indexical properties inform recognition of spoken language (e.g., Goldinger, 1998; Palmeri, et al., 1993; Pisoni, 1997).

In addition to the performance data, listeners trained with a task that emphasised either the linguistic or the indexical properties of the dysarthric signal exhibited similar types of error patterns at the segmental and suprasegmental levels of perceptual processing. Analysis of segmental-level errors revealed no significant difference in the number of syllables the resembled their phonetic target (PSR) between the linguistic and indexical training groups. Thus, even when training encouraged perceptual attention toward indexical elements of the dysarthric signal, listeners gleaned just as much information about the learnable acoustic-phonetic features as listeners trained to attend specifically to the linguistic properties. Furthermore, with a greater number of predicted (IS and DW) versus non-predicted (IW and DS) LBEs for both training groups, the current findings reveal similar segmentation strategies for listeners trained to attend to indexical or linguistic signal properties. Thus, suprasegmental-level errors demonstrated that both training groups adhered to the MSS, which postulates reliance on syllabic stress contrast cues to inform speech segmentation decisions (see Chapter 2, section 2.6.3 for full details). Such a finding may be expected, given the nature of training stimuli—phrases that were designed to specifically draw attention to the alternating syllabic stress of both the training and transcription phrases (see Chapter 2, section 2.5.1). The current findings, therefore, validate those observed in Chapter 4, wherein listeners exposed to the same experimental phrases, under either passive or explicit familiarisation conditions, utilised syllabic stress contrast cues for segmentation of dysarthric speech. It would appear that the experimental phrases utilised in the current programme of research afford robust learning of prosodic cues for segmentation of hypokinetic dysarthric speech.

One difference that arose between training groups was the degree to which the listeners relied on the syllabic stress contrasts cues to inform speech segmentation decisions. The present study indicates that the perceptual strategy of attending to syllabic stress information was most effective for listeners in the indexical training group. This evidence is found in discrepancies in the IS/IW ratios, which reflect strength of adherence to predicted error patterns. Thus, while both experimental groups utilised syllabic stress contrast cues to segment the dysarthric speech, a training task that emphasised properties of the voice enabled listeners to exploit this cue to a greater degree. This finding raises an interesting hypothesis for further testing, that stress patterns may be part of the indexical representation of the acoustic properties of dysarthric speech.

While it is possible that a longer training period would have facilitated more detectable group differences in the learning mechanisms that underlie improved linguistic processing following indexical or linguistic training, significant performance gains relative to baseline intelligibility would suggest the current training paradigm may be sufficient. Furthermore, comparable error patterns at both segmental and suprasegmental levels of perceptual processing for listeners familiarised under different conditions have been reported in Chapter 4. Taken together, error patterns observed with segmental and suprasegmental processing are remarkably similar, regardless of which signal properties are highlighted during training. Accordingly, the current study provides evidence for the interdependence of the learning mechanism responsible for encoding and processing of linguistic and indexical properties. There is now preliminary data to validate the claim that processing of indexical and linguistic properties of the signal may recruit many of the same cognitive-perceptual processes (Nygaard & Pisoni, 1998).

Speculations, however, must be considered relative to the limitations of the study. Listener participants were reasonably accurate (approximately 77% correct) at identifying the word/speaker during the respective training task. This may indicate that task demands were not high enough to facilitate adequate processing of either the linguistic and indexical properties of the signal. Thus, a possible alternative explanation for the comparable perceptual learning outcomes is that the performance data and errors patterns observed were the consequence of familiarisation with the degraded speech, rather than the training task per se. Additional investigations into the influence of indexical information in perceptual learning of dysarthric speech would serve to strengthen the conclusions of the current study.

5.6 CONCLUSION

The present study has provided preliminary evidence that learning to better process dysarthric speech can be sourced from both the linguistic and indexical properties of the signal and yields support for a model of perceptual processing in which learning of indexical properties is encoded and retained alongside the linguistic properties of the signal (e.g., Pisoni, 1997). These observations add to the growing body of literature that challenge long-standing theoretical paradigms that postulate independent processing of such information. Indeed, functional processing of linguistic and indexical information appears to be fundamentally linked.

CHAPTER SIX

Summary, Clinical Implications, Limitations, and Future Directions

6.1 SUMMARY

With insight into the conditions, stimuli, and learning mechanisms that promote improved recognition of dysarthric speech, the programme of research described in this thesis affords preliminary evidence for the development of a theoretical framework that exploits perceptual learning for the management of dysarthria. Background information provided in Chapter 1 acknowledged that reduced intelligibility is a debilitating symptom for individuals with dysarthria and that efficacy data for the use of current behavioural techniques, which are aimed largely at the speaker, is limited. Given that recent definitions of speech intelligibility have highlighted the contributions of both of the speaker and the listener, Chapter 1 raised the possibility that novel approaches to dysarthria management may target learning on behalf of the listener. A review of the literature revealed a large body of evidence for improved understanding of a speech signal that is initially difficult to understand (e.g., synthetic speech, noise-vocoded speech), however, relatively few studies had addressed perceptual learning with the neurologically degraded speech signal. Furthermore, the studies that had reported improved recognition of dysarthric speech were largely clinically based, with limited application to current models of perceptual processing. Chapter 1 called to attention the need for a systematic and theoretically motivated investigation into perceptual learning of dysarthric speech. Thus, the rationale for the exploratory series of studies conducted in this thesis was established.

Chapter 2 detailed a thorough description of the methodology employed across all three phases of the research programme. In brief summary, experimental speech stimuli were collected from three speakers with moderate hypokinetic dysarthria and three neurologically intact control speakers. Across the research phases, 150 healthy listeners participated in a perceptual learning experiment, wherein they were familiarised with stimuli under varying conditions and subsequently transcribed a set of phrases produced by the three speakers with dysarthria. Listener transcripts were then analysed for three primary measures of perceptual processing: intelligibility (PWC), segmental-level processing (PSR), and suprasegmental-level processing (LBE).

Chapter 3 comprised the initial phase of the research programme. Phase one provided strong empirical evidence of improved recognition of dysarthric speech and addressed some of the limitations evident in the existing literature. This was the first study of its kind to directly compare perceptual learning effects for a group of listeners familiarised with dysarthric speech (read passages) to a group of listeners familiarised with the same speech stimuli produced by neurologically intact speakers. A high level of experimental control has enabled more definitive conclusions regarding the effect of familiarisation with dysarthric speech to be established. This study also observed that explicit familiarisation, defined as an experience in which the degraded speech is supplemented with written information, offered performance benefits superior to those afforded by passive familiarisation. Not only was the magnitude of immediate intelligibility gain significantly greater for listeners who received explicit familiarisation, but intelligibility benefits relative to the control group were evident seven days following the initial familiarisation experience. In contrast, listeners who received passive familiarisation did not exhibit intelligibility carry-over at follow-up testing. To date, this is the first large scale study that has directly compared the effects of passive and explicit familiarisation conditions on perceptual learning of dysarthric speech and in addition, documented the longevity of intelligibility benefit with this population over time.

Phase one also investigated possible source(s) of learning associated with the intelligibility improvements observed. Regardless of whether learning conditions were passive or explicit, error pattern analysis indicated improved mapping of acoustic-phonetic information for listeners familiarised with dysarthric speech. Furthermore, these segmental benefits remained robust at follow-up despite significant intelligibility decline. Accordingly, the possibility that learning decay may occur at different rates across different levels of analysis was proposed. Condition discrepancies were evident in lexical segmentation strategies. Listeners who received explicit familiarisation exploited syllabic stress contrasts cues to inform speech segmentation decisions, whereas this acoustic information was largely ignored by listeners who received passive familiarisation. Taken together, the error patterns indicative of cognitive-perceptual processing were taken as evidence that performance differences were more than simply magnitude of benefit. It was, therefore, speculated that the learning afforded by passive familiarisation may be qualitatively different to that which occurred with explicit familiarisation.

The speculation from phase one, that passive and explicit familiarisation conditions may recruit different learning mechanisms, established the rationale for the second phase of the research programme outlined in Chapter 4. Phase two further investigated the notion that the performance differences between passive and explicit learning conditions may be more than magnitude of benefit. In contrast to the use of passage-level stimuli in phase one, listeners in phase two were familiarised with a series of experimental phrases designed to draw attention to the alternating syllabic stress of the transcription phrases. Analysis of speech segmentation error patterns revealed that the condition discrepancy evident in phase one was not robust with the new familiarisation stimuli. Listeners familiarised with experimental phrases utilised syllabic stress contrast cues to segment dysarthric speech, regardless of whether learning conditions were passive or explicit. Thus, Chapter 4 revealed that learning mechanisms were not merely influenced by the presence or absence of written information during familiarisation.

In addition, intelligibility data revealed that performance benefits were only realised when the experimental phrases were supplemented with written information under explicit learning conditions. Furthermore, when conditions were controlled for, performance gains afforded by linguistically-rich passage-level stimuli in phase one were significantly greater than those afforded by the experimental phrases with no sentence-level meaning in phase two. A relationship between the fundamental intelligibility of the familiarisation material and perceptual learning of dysarthric speech was hypothesised. Taken together, the second phase of the research programme identified that perceptual learning, both intelligibility gains and source of learning, may be influenced differentially by the information afforded within the familiarisation procedure.

Collectively, phases one and two revealed that improved recognition of dysarthric speech was only accomplished when linguistic properties of dysarthric speech were highlighted with signal-independent information. The third phase of the research programme acknowledged that the speech signal contains both linguistic and indexical information and accordingly, investigated the role of speaker-specific properties in perceptual learning of neurologically degraded speech. Phase three observed significant intelligibility improvements for listeners trained to attend to indexical signal properties (speaker identification task), and found that these performance gains were comparable to those achieved by listeners trained to attend to linguistic signal properties (word identification task). Improved linguistic processing for listeners trained to attend to indexical features was taken as an indication that indexical properties, as with linguistic properties, of the dysarthric signal are learnable. Furthermore, it was speculated that the processing of both indexical and linguistic information may be intricately linked.

Phase three also observed remarkable similarities in error patterns, indicative of both segmental and suprasegmental processing, following a linguistic or indexical training task. Such findings were considered preliminary evidence that similar cognitive-perceptual mechanisms may be responsible for encoding of linguistic and indexical signal properties. Thus, the final research phase offered empirical validation to support the development of a theoretical model that accounts for the interaction between linguistic and indexical properties as a source of learning in improved recognition of dysarthric speech.

Due to differing research designs and study aims, statistical comparisons between phase three findings and those from the initial two phases were not performed. However, intelligibility gains afforded by the linguistic and indexical training tasks in Chapter 5 were notably smaller than those facilitated by three alternative familiarisation procedures (explicit-phrases, passive-passages, and explicit-passages) employed in Chapter 3 and Chapter 4. The less significant performance gains observed in the final research phase may be explained by the limited feedback (single word or single name) and use of semantically anomalous phrases during training. That an experience involving linguistically-rich passage-level stimuli coupled with supplementary written information enabled the greatest level of intelligibility benefit to be realised suggests that increasing lexical knowledge of the degraded productions may be a critical component in effective exploitation of familiarisation for therapeutic gain. Further research is, however, required test to the efficacy of different familiarisation and training procedures on improved recognition of dysarthric speech (see section 6.3.4).

Overall, the present thesis aimed to offer a theoretically-based perspective on the nature of perceptual learning with dysarthric speech. While moderate hypokinetic dysarthria was employed in the current series of studies, of a number of key findings have theoretical implications that extend across the broader classification of neurologically degraded speech. Firstly, the research found that listeners were able to adapt to a speech signal characterised by segmental and suprasegmental degradation. Secondly, intelligibility improvements were substantially greater when the degraded speech was supplemented with signal-independent information. Thirdly, changes with processing of segmental and suprasegmental information appeared to vary, depending on the familiarisation procedure. Finally, both linguistic and indexical properties informed subsequent recognition of dysarthric speech. While limitations of the research warrant discussion and additional studies are required to complete the picture (see section 6.3), the current body of work offers preliminary evidence for the development of a theoretical framework that may enable the application of perceptual leaning to the management of dysarthria to be realised. Clinical implications are further highlighted in the ensuing section.

6.2 CLINICAL IMPLICATIONS

6.2.1 Novel Practices

This thesis has underscored that significant intelligibility improvements can be achieved for listeners familiarised with dysarthric speech. Given that the primary goal of dysarthria management is to improve speech intelligibility, research that provides evidence of such holds considerable clinical value. That substantial perceptual benefit was observed following a relatively short sample of dysarthric speech suggests that perceptual learning may be an effective and efficient approach for addressing the intelligibility impairments that characterise dysarthric speech. As noted in the review in Chapter 1, the management of dysarthria has traditionally employed behavioural approaches that target the affected speaker. Not only is the evidence base for such treatments limited (e.g., Sellars, et al., 2007), but co-occurring impairments (e.g., reduced motor and cognitive functioning) likely interfere with the individual's capacity to modify behaviour for long-term functional gain (Duffy, 2005). Given this, communication partners (i.e., spouse, family, and/or caregivers) may be in a better position to engage in, and benefit from, focused behavioural rehabilitation programmes. Listener-focused intervention may, therefore, provide an alternative or adjunct approach to existing management options (see Chapter 1, section 1.4.2.5). While a well-researched familiarisation protocol with both familiar and unfamiliar listeners will ultimately be required (see section 6.3), the current programme of research has provided foundational evidence toward the establishment of a treatment approach that targets intelligibility impairments by modifying the perceptual processes of the listener.

6.2.2 Current Practices

The findings of this thesis may also have implications for current approaches to dysarthria management. Significant intelligibility benefits for listeners familiarised with just three short passage readings suggests that, although unstructured, regular contact with patients who exhibit neurologically acquired speech disorders may cause clinicians to better understand dysarthric speech. While the effect of familiarisation upon already familiar listeners, as well as the ability to generalise learning across different forms of dysarthria, both require investigation (see section 6.3), preliminary evidence of the benefit of listeners familiarity (DePaul & Kent, 2000) and learning carry-over (Liss, et al., 2002) has been reported. Thus, the current findings indicate that clinician estimates of baseline intelligibility may be artificially inflated and accordingly, may not reflect the perspective of listeners unfamiliar with dysarthric speech. Ecological validity of the overall description of baseline intelligibility may be increased by augmenting the clinician's score with measures obtained from a naïve listener.

Additionally, the perceptual benefit that transpires with experience may affect the reliability of reporting treatment outcomes of speaker-oriented intervention. Working with a patient on any therapeutic speaking target (e.g., prosody, articulation) may afford significant perceptual benefits for the treating therapist. Thus, as with baseline measures, post-treatment intelligibility scores obtained from the clinician are likely to be exaggerated. Intelligibility improvements may reflect perceptual learning on behalf of the listener rather than significant change to the production of speech. Accurate treatment-related change must, therefore, be measured by a non-treating, and most favourably, listener naïve to dysarthric speech.

Finally, implications can be extended to the broader context of everyday interactions, wherein the ability to express oneself may vary enormously depending on the perceptual experience of the communication partner. Individuals with dysarthria may find that while communication is successful with familiar partners (e.g., therapist, spouse who has adapted to the signal), the same capacity to be understood may not be realised when interacting with unfamiliar members of the general public (e.g., checkout operators, taxi drives, etc). As such, current practice should include patient and family education regarding the variable nature of intelligibility, and empower individuals with supplementary strategies to increase communication success in the presence of unfamiliar listeners.

6.2.3 Other Populations

Finally, while the notion of exploiting perceptual learning for rehabilitative gain has been framed entirely within the context of dysarthria management, the scope of application is potentially much broader. Indeed, treatments that target perceptual processes may be appropriate in any situation in which intelligibility is compromised, including foreign-accented speech, deaf speech, speech processed through cochlear implants, and/or synthesised speech systems. While the source and conditions of learning may be differentially influenced by the nature of the acoustic degradation, the clinical application is the same—improved intelligibility of a speech signal that is initially difficult to understand.

6.3 LIMITATIONS AND FUTURE DIRECTIONS

The studies described in this thesis are limited by a number of factors which have implications for the current findings and future directions for research into perceptual learning of dysarthric speech. While some limitations have been discussed briefly in the experimental chapters, a more comprehensive discussion around these and how such factors may be addressed in future research is highlighted in the subsequent section. Limitations and future directions for research are discussed with regards to the following four methodological variables: (1) speakers, (2) learning source, (3) listeners, and (4) perceptual learning procedures.

6.3.1 Speakers

Perhaps the greatest limitation of the current research was the use of a single form of dysarthria. The perceptual presentation of dysarthria can vary tremendously, not only in type of speech errors and patterns, but also in the degree to which these acoustic disturbances manifest in the speech signal. Traditionally differentiated into seven subtypes given a severity rating (see Chapter 1, section 1.2), the deviant speech characteristics associated with each form of dysarthria may challenge perceptual processing in different ways (see Chapter 1, section 1.3.4). Accordingly, the dysarthrias may differentially influence perceptual learning. While the more general findings generated from the current thesis may be applied across the dysarthrias (see key findings in section 6.1), specific results, particularly those that pertain to underlying cognitive-perceptual processes, should only be attributed to perceptual learning of the dysarthria test case, moderate hypokinetic dysarthria. It is, therefore, acknowledged that similar studies undertaken with different presentations of dysarthric speech would enable a more complete picture of perceptual learning of neurologically degraded speech to be realised. Furthermore, such studies could be extended to investigate if learning effects can carry over from one type and/or severity of dysarthria to another. Whether or not perceptual learning can generalise across the signal characteristics would offer additional understanding into the cognitive-perceptual mechanisms that underlie improved recognition of dysarthric speech.

6.3.2 Learning Source

A second limitation is found in existing attempts to document the source of learning (i.e., “what is learnable?”). The current body of research approached source of learning first in terms of error patterns considered indicative of segmental and suprasegmental processing. Analysis of the PSR data revealed significant changes in phoneme-level perception, regardless of familiarisation conditions (Chapter 3). Thus, it was speculated that listeners learned to better map acoustic-phonetic properties of the dysarthric signal onto existing mental representations of speech sounds. However, the segmental analysis employed in the current studies does not offer information regarding the nature of the phonemic shifts, nor does it reveal which acoustic-phonetic features were responsible for the changes evident at this level of perceptual processing. The LBE analysis provided more specific detail of suprasegmental properties as a source of learning. All three studies demonstrated that syllabic stress properties of the signal provided a source of learning for speech segmentation decisions, although whether these acoustic cues were utilised was dependent on the information emphasised during the familiarisation procedure. Thus, attentional shifts toward syllabic stress contrast information reflect how information afforded by the familiarisation procedure may be leveraged to improve perceptual performance. Additional suprasegmental features of the dysarthric signal as a potential learning source were not examined.

The final phase of the research programme, Chapter 5, continued to document error patterns with processing of segmental and suprasegmental information, but also considered source of learning from a more global perspective of encoding and retaining information afforded within the linguistic and indexical constraints of the signal. The study found that both linguistic and indexical signal properties may provide a source of learning, but did not detail which specific features informed improved processing of dysarthric speech. Thus, the experimental design and analysis employed in the current thesis is limited in its capacity to delineate the locus of learning. If the more general view of perceptual learning is adopted, it could be hypothesised that those production features that are the most consistent and regular will be more learnable. Subsequently, these features would be most salient for improving perceptual performance, relative to those aspects which are inconsistently expressed. By extension, dysarthrias with more consistent signal degradations (e.g., hypokinetic) would be expected to be more amenable to perceptual learning than those with more inherent variability (e.g., hyperkinetic). However, the role of acoustic consistency in perceptual learning remains largely untested. It may very well be that there is also perceptual benefit in exposure to non-systematic acoustic variation, even though the source of benefit could not be attributed to inducing a perceptual remapping. In this case, establishing *expectations of variability* may be the mechanism by which performance is enhanced. Future studies should seek to detail the consistent and inconsistent perceptual features of the different types of dysarthria and employ a comprehensive multi-level analysis that detects which linguistic and indexical cues are most informative in perceptual learning of dysarthric speech.

6.3.3 Listeners

All listener participants were recruited on the basis of minimal or no prior experience with dysarthric speech (naïve listeners) to enable foundational evidence regarding the benefit of a familiarisation procedure to be established. However, in order to extend the clinical applicability of a perceptual learning approach to the management of dysarthria, experimental studies are needed to investigate the influence of listener familiarity (i.e., previous exposure to dysarthric speech) on the benefit of specific familiarisation procedures.

Additionally, all listener participants were required to be aged between 19 and 40 years, and pass a standard pure tone hearing screen. While this inclusion criterion permitted control over age-related variables that may confound learning, it limited the degree to which the finding could be generalised to the older adult population. Ageing is associated with poorer hearing acuity (Morrell, Gordon-Salant, Pearson, Brant, & Fozard, 1996), which has been linked to inferior performance on a variety of language tasks exhibited by older adults (e.g., Sommers, 1997; Wingfield, McCoy, Peelle, Tun, & Cox, 2006). Furthermore, age-related changes have been observed in a number of cognitive operations necessary for language processing, including: decreased attentional resources (McDowd & Shaw, 2000), slower processing of information (Salthouse, 1996), and reduced capacity of working memory (Zacks, Hasher, & Li, 2000). The influence of ageing on perceptual learning of speech has not been widely studied. However, preliminary evidence with a time-compressed signal has suggested that older adults may adapt to degraded speech at a similar rate and magnitude to younger adults, but that maintenance and transfer of this learning may decline with age (Golomb, et al., 2007; Peelle & Wingfield, 2005). To date, no study has examined the interplay of age-related hearing and cognitive decline upon perceptual learning of dysarthric speech. Clinically, recognition of dysarthric speech is important across the lifespan, particularly given that neurological disease (e.g., PD) and injury (e.g., stroke) is more commonly associated with the older population. Spouses of individuals with dysarthria are, therefore, more likely to fall into the older adult category. Accordingly, research that evaluates perceptual learning of dysarthric speech with older adult populations holds significant theoretical and clinical value.

A third limitation pertaining to the listener participants employed in the current study is that the majority were recruited from undergraduate programmes at the University of Canterbury. This, therefore, raises the possibility of an educational/intellect bias, wherein a large number of the participants could perhaps be classified as *optimal* learners. Intellectual factors (e.g., working memory capacity) likely play a role in learning to better recognise difficult speech, and as such, the rapid and substantial adaptation that was observed for listeners familiarised with dysarthric speech in the current programme of research may not reflect the learning capacity of the general population. Future studies that employ broader sampling strategies and additional assessment tests (e.g., working memory and cognitive assessments) will serve to strengthen the present findings. In addition, forthcoming studies should also consider other listener-variables that may interact with perceptual learning, including motivation to learn and previous experience with foreign languages and/or other speech disorders.

6.3.4 Perceptual Learning Procedures

Finally, in order to advance the development of an evidence-based listener-focused approach to the management of dysarthria, future research is required to comprehensively investigate the modifiable variables of the perceptual learning procedure. The present series of studies has shed some light on the nature of the type of familiarisation procedure required to promote improved recognition of dysarthric speech, with the largest and most robust intelligibility scores evidenced following explicit familiarisation to passage-level stimuli. However, a well researched protocol for facilitating the most effective and efficient intelligibility improvements following a specific experience with dysarthric speech has yet to be established. Studies are required to identify the relationship between procedural variables (i.e., stimuli, tasks, amount, and frequency) and intelligibility gain. Furthermore, outcome measures should continue to include word recognition scores, but also encompass additional perceptual measures that pertain to the comprehensibility and naturalness of the signal (see Yorkston, et al., 2006).

6.4 CONCLUSIONS

Perceptual learning of dysarthric speech was identified as an area for further development and research. The current series of studies has provided background information into the nature of perceptual learning with moderate hypokinetic dysarthria. In doing so, this thesis represents an original attempt to investigate the concept of experience-evoked adaptation to a neurologically degraded speech signal and has provided a valuable contribution to this area of research. Furthermore, this thesis has provided a platform for future research in the area of perceptual learning of dysarthric speech. Dysarthria and the associated reduced speech intelligibility continues to be one of the most debilitating symptoms of neurological injury and disease. This research has highlighted that significant intelligibility improvements can be achieved following a short familiarisation experience with dysarthric speech. Although specific listener-focused behavioural approaches have not been considered traditional practice in the management of dysarthria, the current programme of research reveals that this may be a successful avenue to address the intelligibility impairments associated with this speaker population. It is anticipated that ongoing research will enable the full benefit of such an approach to be realised.

APPENDIX A

Familiarisation Stimuli: Passage-Level

The Rainbow Passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow. Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows. Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

Note. The Rainbow Passage comes from Fairbanks, G. (1960). Voice and articulation drillbook (2nd edn). New York: Harper & Row.

APPENDIX B

Familiarisation and Test Stimuli: Phrase-Level

Experimental Phrases

Speech Set One

1. Account for who could knock
2. Admit the gear beyond
3. Amend estate approach
4. And spoke behind her sin
5. Attack became concerned
6. Avoid or beat command
7. Balance clamp and bottle
8. Beside a sunken bat
9. Bush is chosen after
10. Career despite research
11. Connect the beer device
12. Constant willing walker
13. Cool the jar in private
14. Divide across retreat
15. Done with finest handle
16. Had eaten junk and train
17. Frame her seed to answer
18. It's harmful note abounds
19. Increase a grade sedate
20. Indeed a tax ascent
21. Listen final station
22. Mark a single ladder
23. Measure fame with legal
24. Model sad and local
25. Narrow seated member
26. Perceive sustained supplies
27. Rampant boasting captain
28. Resting older earring
29. Rocking modern poster
30. Round and bad for carpet
31. Seat for locking runners
32. Spackle enter broken
33. Submit his cash report
34. Support with dock and cheer
35. Technique but sent result
36. To sort but fear inside

Speech Set Two

1. Address her meeting time
2. Afraid beneath demand
3. Assume to catch control
4. Attend the trend success
5. Award his drain away
6. Bolder ground from justice
7. Cheap control in paper
8. Commit such used advice
9. Confused but roared again
10. Darker painted baskets
11. Define respect instead
12. Distant leaking basement
13. Embark or take her sheet
14. For coke a great defeat
15. Forget the joke below
16. Functions aim his acid
17. Hold a page of fortune
18. Mate denotes a judgment
19. Mistake delight for heat
20. Mode campaign for budget
21. Her owners arm the phone
22. Pick a chain for action
23. Pooling pill or cattle
24. Push her equal culture
25. Remove and name for stake
26. Rode the lamp for testing
27. Rowing father matters
28. Secure but lease apart
29. Signal breakfast pilot
30. Sinking rather tundra
31. Stable wrist and load it
32. Target keeping season
33. Transcend almost betrayed
34. Unless escape can learn
35. Unseen machines agree
36. Vital seats with wonder

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, IL: Aldine.

Adams, S. G. (1997). Hypokinetic dysarthria in Parkinson's disease. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders* (pp. 261-285). New York: Thieme.

Allopena, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements. *Journal of Memory and Language, 38*, 419-439.

Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*, 296-308.

Arboix, A., Marti-Vilalta, J. L., & Garcia, J. H. (1990). Clinical study of 227 patients with lacunar infarcts. *Stroke, 21*, 842-847.

Black, J. W. (1957). Multiple-choice intelligibility tests. *Journal of Speech and Hearing Disorders, 22*, 213-235.

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic charateristics of inadvertently clear speech. *Speech Communication, 14,* 352-337.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaption to non-native speech. *Cognition, 106*, 707-729.

Brown, J., & Carr, T. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1277-1296.

Bunton, K., & Weismer, G. (2001). The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 44*(6), 1215-1228.

Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch and duration. *Journal of Speech and Hearing Disorders, 28*, 221-229.

Canter, G. J. (1965a). Speech characteristics of patients with Parkinson's disease: II. Physiological support for speech. *Journal of Speech and Hearing Disorders, 30,* 44-49.

Canter, G. J. (1965b). Speech characteristics of patients with Parkinson's disease: III. Articulation, diadochokinesis, and overall speech adequacy. *Journal of Speech and Hearing Disorders, 30*, 217-224.

Chenery, H. J., Murdoch, B. E., & Ingram, J. C. L. (1988). Studies in Parkinson's disease: I. Perceptual speech analysis. *Australian Journal of Human Communication Disorders, 16*, 17-29.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaption to foreign-accented English. *Journal of the Acoustical Society of America, 116*, 3647-3658.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 37,* 463-480.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics*, *56*, 624-636.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37*, 463-480.

Costanzo, F. S., Markel, N. N., & Costanzo, P. R. (1989). Voice quality profile and perceived emotion. *Journal of Counseling Psychology*, *16*, 267-270.

Creelman, C. D. (1957). The case of the unknown talker. *Journal of the Acoustical Society of America*, *29*, 655.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218-236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113-121.

D'Innocenzo, J., Tjaden, K., & Greenman, G. (2006). Intelligibility in dysarthria: Effects of listener familiarity and speaking condition. *Clinical Linguistics and Phonetics, 20*(9), 659-675.

Dahan, D., & Mead, R. L. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 704-728.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Differential diagnosis patterns of dysarthria. *Journal of Speech and Hearing Research, 12*, 246-269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 262-496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor Speech Disorders*. Philadelphia: W.B. Saunders.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*, 132-147.

Davis, M. H., Johnsrude, I. S., Herrvais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222-241.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218-244.

Deane, K., Whurr, R., Playford, E. D., Ben-Shlomo, Y., & Clarke, C. E. (2009). Speech and language therapy for dysarthria in Parkinson's disease: a comparison of techniques (Review). *Cochrane Database of Systematic Reviews*, (1), CD002814. doi: 10.1002/14651858

DePaul, R., & Kent, R. D. (2000). A longitudinal case study of ALS: Effects of listeners familiarity and proficiency on intelligibility judgements. *American Journal of Speech-Language Pathology, 9*, 230-240.

Dongilli, P. (1994). Semantic context and speech intelligibility. In J. Till, K. M. Yorkston & D. R. Beukelman (Eds.), *Motor speech disorders: Advances in assessment and treatment*. Baltimore, MD: Paul H. Brookes.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management* (2nd ed.). St. Louis, MS: Elsevier Mosby.

Duffy, J. R. (2007). History, current practice, and future trends and goals. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent*. San Diego, CA: Plural.

Duffy, J. R., & Yorkston, K. M. (2003). Medical interventions for spasmodic dysphonia and some related conditions: A systematic review. *Journal of Medical Speech-Language Pathology, 11*, 9-58.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914-927.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics, 67*(2), 224-238.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America, 119,* 1950-1953.

Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). New York, NY: Harper & Row.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature (London), 425*, 614-616.

Forrest, K., Weismer, G., & Turner, G. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *Journal of the Acoustical Society of America, 85*, 2608-2622.

Francis, A. L., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics, 62*, 1668-1680.

Francis, A. L., & Nusbaum, H. C. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, and Psychophysics*, *71*(6), 1360-1374.

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic-phonetic representation of synthetic speech. *Journal of Speech, Language, and Hearing Research, 50*, 1445-1465.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*, 412-429.

Garcia, J. M., & Cannito, M. P. (1996). Top down influences on the intelligibility of a dysarthric speaker: Addition of natural gestures and situational context. In D. Robin, K. M. Yorkston & D. R. Beukelman (Eds.), *Disorders of motor speech* (pp. 67-87). Baltimore, MD: Paul H. Brookes.

Garcia, J. M., & Dagenais, P. A. (1998). Dysarthria sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language and Hearing Research, 41*, 1282-1293.

Garnes, S., & Bond, Z. S. (1976). The relationship between acoustic information and semantic expectation. *Phonologica*, 285-293.

Golbe, L. I., & Langston, J. W. (1993). The etiology of Parkinson's disease: New directions for research. In J. Jankovic & E. Tolosa (Eds.), *Parkinson's Disease and Movement Disorders* (pp. 93-101). Baltimore, MD: Williams & Wilkins.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.

Goldinger, S. D. (1998). Echoes of echoes?: an episodic theory of lexical access. *Psychological Review, 105*, 251-279.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 152-162.

Goldstone, R. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaption to time-compressed speech. *Journal of the Acoustical Society of America*, *121*(3), 1701-1708.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 421-433.

Grewel, F. (1957). Dysarthria in post-encephalitic Parkinsonism. *Acta Psychiatrica Scandinavica, 32*, 440-449.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, *102*, 655-658.

Halle, M. (1985). *Speculations about the representation of words in memory*. New York, NY: Academic Press.

Hammen, V. L., Yorkston, K. M., & Dowden, P. A. (1991). Index of contextual intelligibility: Impact of semantic context in dysarthria. In C. Moore, K. M. Yorkston & D. R. Beukelman (Eds.), *Dysarthria and apraxia of speech: Perspectives on management* (pp. 43-53). Baltimore, MD: Paul H. Brookes.

Hanson, E. K., Yorkston, K. M., & Beukelman, D. R. (2004). Speech supplementation techniques for dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology, 12*(2), 9-29.

Harteliusa, L., Runmarkerb, B., & Andersenb, O. (2000). Prevalence and characteristics of dysarthria in a multiple-sclerosis incidence cohort: Relation to neurological data. *Folia Phoniatrica et Logopaedica, 52*(4), 160-177.

Hervais-Adelman, A., Davis, M. H., Johnsonrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise-vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 460-474.

Hervais-Adelman, A., Davis, M. H., Taylor, K., Johnsrude, I. S., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 283-295.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099-3111.

Hood, J. D., & Poole, J. P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *Audiology, 19*, 435-455.

Hoover, J. R., Reichle, J., Van Tasell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo versus votrax. *Journal of Speech and Hearing Research, 30,* 425-431.

Howes, D. (1957). On the relation between the intelligibility of speech as a function of the context of the test materials. *Journal of the Acoustical Society of America, 29*, 296-305.

Hustad, K. C. (1996). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatrica et Logopaedica, 58*, 217-228.

Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research, 44*(3), 407-509.

Hustad, K. C., Beukelman, D. R., & Yorkston, K. M. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language, 19*, 291-302.

Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12*, 198-206.

Hustad, K. C., & Weismer, G. (2007). Interventions to improve intelligibility and communicative success for speakers with dysarthria In G. Weismer (Ed.), *Motor Speech Disorders: Essays for Ray Kent*. San Diego, CA: Plural Publishing.

Jankovic, J. (2008). Parkinson’s disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery, and Psychiatry, 79*, 368-376.

Jankovic, J., & Marsden, D. (1993). Therapeutic strategies in Parkinson's disease. In J. Jankovic & E. Tolosa (Eds.), *Parkinson's Disease and Movement Disorders* (pp. 115-144). Baltimore, MD: Williams & Williams.

Joos, M. A. (1948). Acoustic phonetics. *Language Monographs, 23*(23), 136.

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing, 23*, 2-40.

Kennedy, M. R. T., Strand, E. A., & Yorkston, K. M. (1994). Selected acoustic changes in the verbal repairs of dysarthric speakers. *Journal of Medical Speech-Language Pathology, 2*, 263-279.

Kirk, K. I., Pisoni, D. B., & Miyamto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research, 40*, 1395-1405.

Klasner, E. R., & Yorkston, K. M. (2005). Speech intelligibility in ALS and HD dysarthria: The everyday listener's perspective. *Journal of Medical Speech-Language Pathology, 13*(2), 127-139.

Knuth. (1998). *The Art of Computer Programming* (3rd ed. Vol. 2). Boston: Addison-Wesley.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science, 16*, 617-625.

Koul, R., & Hester, K. (2006). Effects of repeated listening experiences on the recognition of synthetic speech by individuals with severe intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 49*, 47-57.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141-178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262-268.

Levi, S. V., & Pisoni, D. B. (2007). Indexical and linguistic channels in speech perception: Some effects of voiceovers on advertising outcomes. In T. M. Lowrey (Ed.), *Psycholinguistics Phenomena in Marketing Communications* (pp. 203-219): Lawrence Erlbaum Associates.

Lindbolm, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Liss, J. M. (2007). The role of speech perception in motor speech disorders. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 187-219). San Diego, CA: Plural.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 112*(6), 3022-3030.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America, 104*(4), 2457-2466.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C. A., & Edwards, B. (2000a). Lexical boundary error analysis in hypokinetic dysarthria. *Journal of the Acoustical Society of America, 107*(6), 3415-3424.

Lockhart, R. S. (2002). Levels of processing, transfer-appropriate processing, and the concept of robust encoding. *Memory, 10*, 397-403.

Loebach, J. L., Bent, T., & Pisoni, D. B. (2008). Multiple routes to the perceptual learning of speech. *Journal of the Acoustical Society of America, 124*(1), 552-561.

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology, 36*(1), 224-234.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Logemann, J. A., Boshes, B., & Fisher, H. B. (1973). The steps in the degeneration of speech and voice control in Parkinson's disease. In J. Siegfried (Ed.), *Parkinson's diseases: Rigidity, akensia, behavior* (pp. 101-112). Vienna, Austria: Hans Huber.

Logemann, J. A., Fisher, H. B., Boshes, B., & al, e. (1978). Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson's patients. *Journal of Speech and Hearing Disorders, 43*, 348-352.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighbourhood activation, and PARSYN. *Perception and Psychophysics, 62*, 615-625.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*, 565-581.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighbourhood activation model. *Ear and Hearing, 19*, 1-36.

Ludlow, C. L., & Bassich, C. J. (1983). *The results of acoustic and perceptual assessment of two types of dysarthria*. San Diego, CA: College-Hill.

Ludlow, C. L., & Bassich, C. J. (1984). Relationships between perceptual ratings and acoustic measures of hypokinetic speech. In M. R. McNeil, J. C. Rosenbek & A. Aronson (Eds.), *The Dysarthrias* (pp. 163-195). San Diego, CA: College Hill.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 676-684.

Massaro, D. W. (2001). Speech Perception. In N. M. Smelser, P. B. Baltes & W. Kintsch (Eds.), *International Encylopedia of School and Behavioral Sciences* (pp. 14870-14875). Amsterdam, The Netherlands: Elseview.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics, 70*(7), 1235-1242.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477-500.

McAuliffe, M. J., Borrie, S. A., Good, P. V., & Hughes, L. (2010). Consideration of the listener in the assessment and treatment of dysarthria. *ACQuiring knowledge in Speech, Language and Hearing, 12*, 16-19.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McDowd, J. M., & Shaw, R. J. (2000). Attention and aging: A functional perspective. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (2 ed., pp. 221-292.). Mahwah, NJ: Lawrence Erlbaum Associates.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners *Journal of Speech and Hearing Research, 26*, 451-458.

McGettigan, C., Rosen, S., & Scott, S. K. (2008). Investigating the perception of noise-vocoded speech - an individual differences approach. *Journal of the Acoustical Society of America, 123*(5), 3330-3330.

McQueen, J. M., & Cutler, A. (1998). Segmentation of continuous speech recognition and the contents of the vocabulary. *Journal of Memory and Language, 39*, 21-46.

McQueen, J. M., & Cutler, A. (2001a). Spoken word access processes: An introduction. *Language and Cogntive Processes, 10*(3), 469-490.

McQueen, J. M., & Cutler, A. (2001b). Spoken word acess processes: An introduction. *Language and Cognitive Processes, 10*(3), 309-331.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*, 1113-1126.

McQueen, J. M., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proceedings of ISCA Workshop on Plasticity in Speech Perception*, 233-236.

Mehler, J., Sebastian-Galles, N., Altmann, G. T. M., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Sciences*, *682*, 272-282.

Melo, T. P., Bogousslavsky, J., van Melle, G., & Regli, F. (1992). Pure motor stroke: a reappraisal. *Neurology, 42*, 789-798.

Milenkovic, P. (2001). TF32: Time-frequency analysis for 32-bit windows [Computer software]. Madison: Wisconsin.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology, 41*, 329-335.

Miller, J. L. (1981). Effects of speaking rate on segmental sidtinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occuring information on the perception of stop consonant and semi-vowel. *Perception and Psychophysics, 25*(6), 457-465.

Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., & Fozard, J. L. (1996). Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level. *Journal of the Acoustical Society of America, 100*, 1949-1967.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics, 47*, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics, 34*, 202-240.

Murry, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*(1097-1108).

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62*, 714-719.

Nearey, T. (1978). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85*, 2088-2113.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America, 95*, 1085-1099.

Nooteboom, S. G. (1981). Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics, 9*, 407-424.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204-238.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39-57.

Nusbaum, H. C., & Goodman, J. (1994). Learning to hear speech as spoken language. In H. C. Nusbaum & J. Goodmann (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 299-338). Cambridge, MA: MIT Press.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern Recognition by*

Humans and Machines: Volume 1. Speech Perception. (pp. 113-157). New York: Academic Proess.

Nygaard, L. C. (2008). Perceptual integration of linguistic and nonlinguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. Malden, MA: Blackwell Publishing.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60*, 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42-46.

Pallier, C., Sebastian-Galles, N., Dupoux, E., & Christophe, A. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition, 26*, 844-851.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 915-930.

Parkinson, J. (1817). *An Essay on the Shaking Palsy*. London: Sherwood, Neely and Jones.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of the Acoustical Society of America, 31*, 1315-1330.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*, 693-703.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-32). San Diego, CA: Academic Press.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In H. C. Nusbaum & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

Poewe, W. (1993). L-dopa in Parkinson's disease: Mechanisms of action and pathophysiology of late failure. In J. Jankovic & E. Tolosa (Eds.), *Parkinson's Disease and Movement Disorders* (pp. 103-111). Baltimore, MD: Williams & Williams.

Rajaram, S., Srinivas, K., & Roediger, H. (1998). A transfer-appropriate processing account of context effects in word-fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 993-1004.

Reynolds, M. E., Isaacs-Duvall, C., & Haddox, M. L. (2002). A comparison of learning curves in natural and synthesized speech comprehension. *Journal of Speech, Language, and Hearing Research, 45*, 802-810.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*, 3629-3636.

Rosenbek, J. C., & LaPointe, L. L. (1985). The dysarthrias: Description, diagnosis, and treatment In D. Johns (Ed.), *Clinical Management of neurogenic communication disorders*. Boston: Little Brown

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33*(3), 441-452.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474-494.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207-1218.

Sandyk, R. (1995). Resolution of dysarthria in multiple sclerosis by treatment with weak electromagnetic fields. *International Journal of Neuroscience, 83*, 81-92.

Sapir, S., Pawlas, A. A., Ramig, L. O., Countryman, S., O'Brien, C., Hoehn, M. M., et al. (2001). Voice and speech abnormalities in Parkinson's disease: Relation to severity of motor impairment, duration of disease, medication, depression, gender, and age. *Journal of Medical Speech-Language Pathology, 9*, 213-226.

Sarno, M. T., Buonaguro, A., & Levita, E. (1986). Characteristics of verbal impairment in closed head injured patients. *Archives of Physical Medicine & Rehabilitation, 67*, 400-405.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123-148.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395-408.

Sebastian-Galles, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception and Psychophysics, 62*, 834-842.

Sellars, C., Hughes, T., & Langhorne, P. (2007). Speech and language therapy for dysarthria due to non-progressive brain damage. *Cochrane Database of Systematic Reviews*, (3), CD002088. doi: 10.1002/14651858

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primary temporal cues. *Science, 62*(4), 834-842.

Sommers, M. S. (1997). Stimulus variability and spoken word recognition: II. The effects of age and hearing impairment. *Journal of the Acoustical Society of America, 101*, 2278-2288.

Spencer, K. A., Yorkston, K. M., & Duffy, J. R. (2003). Behavioral management of respiratory/phonatory dysfunction from dysarthria: A flowchart for guidance in clinical decision making. *Journal of Medical Speech-Language Pathology, 11*, 39-61.

Spitzer, S. M., Liss, J. M., Caviness, J. N., & Adler, C. (2000). An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech. *Journal of Medical Speech-Language Pathology, 8*, 285-293.

Stacey, P. C., & Summerfield, A. Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America, 121*, 2923-2935.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research, 18*, 162-171.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after effect. *Perception and Psychophysics, 35*, 203-213.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition, priming and word identification. *Psychonomic Bulletin & Review, 2,* 339-363.

Theodoros, D. G., Murdoch, B. E., & Goozee, J. V. (2001). Dysarthria following traumatic brain injury: Incidence, recovery, and perceptual features. In B. E. Murdoch & D. G. Theodoros (Eds.), *Traumatic brain injury: Associated speech, language, and swallowing disorders* (pp. 27-51). San Diego, CA: Singular.

Tikofsky, R. S., & Tikofsky, R. P. (1964). Intelligibility as a measure of dysarthric speech. *Journal of Speech and Hearing Research, 7*, 325-333.

Tjaden, K. K., & Liss, J. M. (1995a). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics and Phonetics, 9*(2), 139-154.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition - Patterns and parameters 1. Recognition of backward voices. *Journal of Phonetics, 13,* 19-38.

Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition - Patterns and parameters 2. Recognition of rate-altered voices. *Journal of Phonetics, 13,* 39-52.

Vogel, D., & Miller, L. (1991). A top-down approach to treatment of dysarthric speech. In D. Vogel & M. Cannito (Eds.), *Treating disordered speech motor control* (pp. 87-109). Austin, TX: Pro-Ed.

Volkmann, A. (1858). Ueber den Einfluss der Uebung auf das Erkennen räumlicher Distanzen. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe, 10,* 38-69.

Warlow, C. P., Dennis, M. S., van Gijn, J., Hankey, G. J., Sandercock, P. A. G., Bamford, J. G., et al. (1996). *Stroke: a practical guide to management*. Oxford: Blackwell Scientific.

Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, *41*, 262-275.

Warren, P., & Marslen-Wilson, W. D. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics, 43,* 21-30.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American Journal 223*, 30-36.

Watkins, A. (1981). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America, 90,* 2942-2955.

Weill, S. A. (2001). Foreign-accented speech: Encoding and generalization. *Journal of the Acoustical Society of America, 109,* 2473 (A).

Weiner, W. J., & Lang, A. E. (1989). *Movement disorders: A comprehensive survey*. Mount Kiskon, NY: Futura.

Weismer, G. (1984). Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences. *Seminars in Speech and Language, 5,* 293-314.

Weismer, G. (2006). Philosophy of research in motor speech disorders. *Clinical Linguistics and Phonetics, 20,* 315-349.

Weismer, G., & Kim, Y. (2010). Classification and Taxonomy of Motor Speech Disorders: What are the Issues? In B. Maassen & P. vanLieshout (Eds.), *Speech Motor Control:*

New developments in basic and applied research (pp. 229-242). Oxford: Oxford University Press.

Wijesekera, L. C., & Leigh, P. N. (2009). Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases, 4*, 1-22.

Wingfield, A., McCoy, S. L., Peelle, J. E., Tun, P. A., & Cox, L. C. (2006). Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *Journal of the Academy of Audiology, 17*, 487-497.

Yohan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older adults. *Psychology and Aging, 15*(88-99).

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*, 499-512.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of intelligibility of dysarthric speech*: Tigard, OR: CC Publications.

Yorkston, K. M., & Beukelman, D. R. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. Berry (Ed.), *Clinical dysarthria* (pp. 155-164). Austin, TX: Pro-Ed.

Yorkston, K. M., Beukelman, D. R., & Bell, K. R. (1988). *Clinical management of dysarthric speakers*. Austin, TX: ProEd.

Yorkston, K. M., Beukelman, D. R., & Hakel, M. (1996). Speech Intelligibility Test for Windows. Lincoln, NE: Communication Disorders Software.

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Bell, K. R. (1999). *Management of motor speech disorders in children and adults*. Austin, TX: Pro-Ed.

Yorkston, K. M., Dowden, P. A., & Beukelman, D. R. (1992). Intelligibility as a tool in the clinical management of dysarthric speakers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement and management* (pp. 265-286). Amsterdam, The Netherlands: John Benjamins.

Yorkston, K. M., Hakel, M., Beukelman, D. R., & Fager, S. (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic reveiw. *Journal of Medical Speech-Language Pathology, 15*(2), 11-36.

Yorkston, K. M., Hammen, V. L., Beukelman, D. R., & Traynor, C. D. (1990). The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders, 55*, 550-561.

Yorkston, K. M., Honsinger, M. J., Mitsuda, P. M., & Hammen, V. L. (1989). The relationship between speech and swallowing disorders in head-injured patients. *Journal of Head Trauma Rehabilitation, 4*, 1-16.

Yorkston, K. M., Spencer, K., Duffy, J. R., Beukelman, D. R., Golper, L. A., Miller, R., et al. (2001). Evidence-based practice guidelines for dysarthria: Management of velopharyngeal function. *Journal of Medical Speech-Language Pathology, 9*(4), 257-273.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (2006). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*, 55-66.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (pp. 293-357.). Mahwah, NJ: Lawrence Erlbaum Associates.

Zwitserlood, P. (1989). The locus of the effects of the sentential-semantic context in spoken word processing. *Cognition*, *32*, 25-64.